

# Coarse Graining of Agent-Based Models and Spatio-Temporal Modeling of Spreading Processes

**Dissertation**

zur Erlangung des akademischen Grades eines  
Doktors der Naturwissenschaften

vorgelegt von

**Johannes Zonker**

Fachbereich Mathematik und Informatik  
Freie Universität Berlin

Berlin, März 2023

**Betreuer:** Prof. Dr. Christof Schütte

**Erstgutachter:** Prof. Dr. Christof Schütte

**Zweitgutachter:** Prof. Dr. Wilhelm Huisinga

**Tag der Disputation:** 9. November 2023

**Selbstständigkeitserklärung**

Ich erkläre gegenüber der Freien Universität Berlin, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Die vorliegende Arbeit ist frei von Plagiaten. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht. Diese Dissertation wurde in gleicher oder ähnlicher Form noch in keinem früheren Promotionsverfahren eingereicht.

Mit einer Prüfung meiner Arbeit durch ein Plagiatsprüfungsprogramm erkläre ich mich einverstanden.

Berlin, 23.03.2023

---

Johannes Zonker

# Acknowledgements

It is my great pleasure to thank the many people who have supported me in my research for this thesis and in general. First of all, I would like to thank Christof Schütte for introducing me to this interesting research area, for giving me the opportunity to write this thesis, and for his constant support over the years. A very special thanks goes to Nataša Djurdjevac Conrad for many fruitful collaborations and discussions as well as for her guidance and encouragement. I would also like to thank Stefanie Winkelmann for helpful scientific advice and the time of collaboration. Many thanks also to Péter Koltai and Marcus Weber for insightful scientific discussions.

This thesis would not have been possible without the interdisciplinary collaboration in the various projects I have worked on. Thus, in alphabetical order, I would like to thank Fabian Becker, Raphael A. Eser, Daniel Fürstenau, Ana Grabundzija, Cecilia Padilla-Iglesias, Martin Park, Wolfram Schier, and Brigitta Schütt for all their contributions, from providing data sets to the many interesting and fruitful discussions during the development and analysis of the models.

I would also like to thank the Biocomputing Group at the Freie Universität Berlin and the Computational Humanities Group at the Zuse Institute Berlin for the pleasant working atmosphere, and the Computational Systems Biology Group at the Zuse Institute Berlin for the scientific discussions in our joint group seminars. Special thanks to Jan-Hendrik Niemann, András Tóbiás, and Niklas Wulkow for proofreading parts of the manuscript.

Many thanks to all the musicians who performed with me in the last years and to my singing teacher. For me there is no greater joy than music and it is the foundation for everything in my life. Thanks to all my friends for making my life enjoyable and keeping me motivated. Last but not least I would like to thank my parents for supporting every decision I have made in my life and encouraging me every step of the way.

I am grateful for the funding of my work by the German Cluster of Excellence MATH+, the German Research Foundation, the Einstein Foundation Berlin, the Berlin Mathematical School, the Free University Berlin and the Zuse Institute Berlin.

# Contents

|  |           |
|--|-----------|
| <b>Introduction</b>  | <b>1</b>  |
| <b>1 Mathematical Formalization and Simulation of Agent-Based Models</b> | <b>4</b>  |
| 1.1 Markov Processes   | 4         |
| 1.1.1 General Concepts   | 5         |
| 1.1.2 Markov Diffusion Processes   | 13        |
| 1.1.3 Markov Jump Processes  | 18        |
| 1.1.4 Piecewise-Deterministic Markov Processes                           | 22        |
| 1.1.5 Jump Diffusion Processes   | 25        |
| 1.2 Mathematical Formalization of Agent-Based Models                     | 27        |
| 1.2.1 Definition of Agent-Based Models                                   | 28        |
| 1.2.2 Formalization of ABMs as Markov Processes                          | 31        |
| 1.2.3 Guiding Example  | 36        |
| 1.2.4 Self Organisation of Agents According to Local Information         | 38        |
| 1.3 Simulation of Agent-Based Models                                     | 45        |
| 1.3.1 Simulation of Diffusion Processes                                  | 46        |
| 1.3.2 Simulation of Jump Processes                                       | 48        |
| 1.3.3 Combined Simulation Approach                                       | 52        |
| <b>2 Coarse Graining of Agent-Based Models</b>                           | <b>64</b> |
| 2.1 Markov State Model Theory  | 67        |
| 2.1.1 Galerkin Projection Approach                                       | 67        |
| 2.1.2 Metastability  | 71        |
| 2.1.3 Core Set Approach  | 73        |
| 2.2 Stochastic Metapopulation Model                                      | 76        |
| 2.2.1 SMM Definition   | 77        |
| 2.2.2 Galerkin Projection Approach for ABMs                              | 78        |
| 2.2.3 Projection of the Spatial Dynamics                                 | 81        |
| 2.2.4 Projection of the Adoption Dynamics                                | 84        |
| 2.2.5 Application to the Guiding example                                 | 93        |
| 2.3 Piecewise-Deterministic Metapopulation Model                         | 95        |
| 2.3.1 ODE Approximation for the PDMM                                     | 96        |
| 2.3.2 Application to the Guiding Example                                 | 97        |
| 2.4 Simulation and Computational Efficiency                              | 99        |
| 2.4.1 Simulation of the SMM  | 99        |
| 2.4.2 Simulation of the PDMM   | 100       |
| 2.4.3 Effort Comparison  | 101       |

|          |   |            |
|----------|---|------------|
| <b>3</b> | <b>Modeling of Spreading Processes</b>                          | <b>102</b> |
| 3.1      | The Spreading of the Woolly Sheep in Ancient Times . . . . .    | 104        |
| 3.1.1    | ABM for the Spreading of the Woolly Sheep . . . . .             | 107        |
| 3.1.2    | PDMM Approximation . . . . .                                    | 117        |
| 3.1.3    | Comparison of both Approaches . . . . .                         | 120        |
| 3.2      | Cultural Evolution of Hunter-Gatherer Societies . . . . .       | 126        |
| 3.2.1    | ABM for Mobility and Cultural Evolution of Hunter-Gatherers . . | 127        |
| 3.2.2    | Computational Analysis . . . . .                                | 140        |
| 3.3      | Epidemic Spreading . . . . .                                    | 153        |
| 3.3.1    | General Model Assumptions for Epidemic Spreading . . . . .      | 154        |
| 3.3.2    | PDMM for Epidemic Spreading . . . . .                           | 156        |
| 3.3.3    | Discussion of ABM Approaches and Model Reduction . . . . .      | 167        |
|          | <b>Conclusion</b>   | <b>170</b> |
|          | <b>Bibliography</b>   | <b>172</b> |

*"On the beach at night alone,  
As the old mother sways her to and fro singing her husky song,  
As I watch the bright stars shining, I think a thought of the clef of the universes and of  
the future.*

*A vast similitude interlocks all,  
All spheres, grown, ungrown, small, large, suns, moons, planets,  
All distances of place however wide,  
All distances of time, all inanimate forms,  
All souls, all living bodies, though they be ever so different, or in different worlds,  
All gaseous, watery, vegetable, mineral processes, the fishes, the brutes,  
All nations, colors, barbarisms, civilisations, languages,  
All identities that have existed or may exist on this globe, or any globe,  
All lives and deaths, all of the past, present, future,  
This vast similitude spans them, and always has spann'd  
And shall forever span them and compactly hold and enclose them."*

**Walt Whitman, On the Beach at Night Alone**

# Introduction

The recent events surrounding the COVID-19 pandemic have highlighted the importance of understanding spreading processes within human societies. Not only the epidemic spreading itself, but also the process of informing the public and the spreading of misinformation have challenged scientists and policymakers [1]. Mathematical models are useful for studying these processes and for discussing the similarities and differences in their dynamics. By integrating real-world data, short-term predictions can be made to support the decision-making process [2]. Depending on the scale of the system, different approaches are used to model spreading processes, such as compartmental models based on ordinary differential equations for a macroscopic description of the process, metapopulation models on the mesoscale or the agent-based model (ABM) on the micro level. The choice of model type for an application usually depends on the available data that can be used to parameterize and validate the model, and on the precise research question. Questions that deal with details of the real-world system under consideration, such as the effectiveness of specific interventions, often require a microscopic description of the process through an ABM for meaningful discussion. The amount of detail that can be incorporated into ABMs is one reason why they are a popular choice in many disciplines such as ecology [3], sociology [4], economics [5] and epidemiology [6] for modeling emergent phenomena and processes based on interactions. In recent years, agent-based models have been increasingly used in the field of archaeology to test and illustrate hypotheses about the past [7]. While an increasing amount of data is available on contemporary social processes, data on prehistoric processes are usually scarce and uncertain. Therefore, the application of data-driven statistical approaches is generally infeasible. In this situation, ABMs can be used to fill the gaps in the data through rational modeling assumptions, not with the goal of reconstructing the past, but to provide plausible scenarios that could have happened [8].

Much work remains to be done to find a general framework that unifies the different approaches to ABM used in the literature. Since ABMs are often the result of interdisciplinary collaborations, the models are usually formulated in a mixture of the languages of the fields involved. Thus, ABM formulations are often descriptive rather than mathematically explicit. In this thesis, we will explore a mathematical viewpoint on ABMs that can be formalized as Markov processes. The class of Markov processes is general enough to include particle-based ABMs [9] as well as ABMs based on social networks [10, 11] in one framework. The formalization of ABMs as Markov processes also provides a way to address typical challenges in agent-based modeling. ABMs are comparatively realistic models of dynamics arising from interactions, but they are often very expensive to simulate. Especially for large-scale agent systems with randomness,

generating enough trajectories for statistical analysis or parameter optimization is generally not feasible. However, for Markov processes we can utilize and further develop established simulation algorithms [12] for a gain in computational efficiency and also apply model reduction techniques for metastable dynamics [13] that greatly reduce the simulation effort, though at the cost of losing some details of the original process. The model reduction framework for ABMs that will be established in this thesis builds on the Markov state model theory [13] and provides an explicit connection between metastable ABMs and a mesoscopic formulation as a metapopulation model.

In addition to these more theoretical contributions, we will also examine several models of spreading processes in real-world systems and illustrate how the theoretical results can be applied to them. We will discuss the details of the construction of the models as well as the integration of real-world data and highlight the different challenges for the prehistoric and contemporary context. In the computational analysis of the models one focus is the identification of metastable sets through the application of clustering methods. We will also discuss a novel approach to identify spatio-temporal clusters despite strong perturbations that are generated by the dynamics of the model.

## Outline of the Thesis

This thesis is composed of three chapters. The topic of Chapter 1 is the mathematical formalization and simulation of agent-based models. Chapter 2 focuses on the topic of model reduction based on the Markov state model framework. In the final chapter, the concepts of the previous two chapters are applied to the modeling of spreading processes in real-world systems.

Chapter 1 begins with an overview about the theory of stochastic processes in which the important mathematical objects and terms that are used throughout the thesis are briefly introduced. Afterwards, we proceed with the definition of a framework for the formalization of ABMs as Markov processes. In this section also the basic components of ABMs for spreading processes are introduced within this framework. The chapter concludes with a section about the details of the simulation algorithms that are used in this thesis. All concepts of the first chapter are illustrated with numerous examples, one of them being an ABM for the self-organisation of agents according to local information about a target distribution.

The first section of Chapter 2 is an introduction to the Markov state model framework for the model reduction of metastable dynamics. In the following section this approach is applied to ABMs and an explicit relation between the microscopic and mesoscopic model parameters is established. The model reduction of ABMs to a stochastic and subsequently a piecewise-deterministic metapopulation model is illustrated with a simple example for which we discuss the approximation quality as well as the gain in computational efficiency.

In Chapter 3 several models for prehistoric and contemporary spreading processes that are based on real-world data are explored. The first section features the construction of an ABM for the spreading of the wool-bearing sheep, which is an example for an important innovation in human history. This model is a result of an interdisciplinary cooperation with researchers from the former German excellence cluster TOPOI. In Section 3.2 the ABM for innovation spreading is generalized to a conceptual model for the mobility and cultural evolution of hunter-gatherer societies. Both applications are in the prehistoric context and it is illustrated how such models can be formulated and parametrized even though there is a lack of data about the original processes. Finally, Section 3.3 covers the modeling of epidemic spreading, especially in the context of the COVID-19 pandemic. Several modeling approaches are discussed in this section as well as the details of a metapopulation model that includes local and global intervention strategies in the model assumptions and is parametrized according to the results of recent studies.

## Relation to Previous Publications

The results of this thesis have partially been published in [14–16] or are about to be published in [17]<sup>1</sup>. The event-based simulation algorithm for ABMs was presented in [15] for a less general case, the formalization of ABMs as well as the model reduction framework and the metapopulation model for the spreading of COVID-19 were presented in [16], the ABM for the spreading of the woolly sheep was featured in [14, 15] and the ABM for the mobility and cultural evolution of hunter-gatherer societies is also the topic of [17]. Figures 1.10, 2.1–2.6 and 3.24–3.28 have been originally designed for [16], Figures 1.15, 1.16 and 3.2 have been originally designed for [15], Figure 3.1 has been originally designed for [14] and Figures 3.12–3.14, 3.16, 3.18, 3.22 have been originally designed for [17]. All other figures have been designed for this thesis by the author.

## Model Code Availability

All models have been implemented with MATLAB and the scripts are available in the repository <https://git.zib.de/bzfonke>.

---

<sup>1</sup>The article "Insights into drivers of mobility and cultural dynamics of African hunter-gatherers over the past 120,000 years" is now published. The citation has been updated accordingly.

# 1 Mathematical Formalization and Simulation of Agent-Based Models

In this chapter we provide a framework for formulating an agent-based model (ABM) with explicit mathematical equations and objects, which often allows formal analysis as well as the application of established techniques for efficient simulation, computational analysis and model reduction.

For the modeling of real-world processes the introduction of randomness to the model is often essential to match the characteristics of the process and to include the fact that even with a lot of input data we are always working with incomplete information. Especially when it comes to the modeling of human interactions, there are no deterministic laws that human decision-making always follows (e.g., due to free will [18]). However, we can often assume that there are many independent factors that influence a decision and that therefore the law of large numbers and the central limit theorem hold. For these reasons, we make assumptions about the characteristics of the randomness that is necessary to formulate our agent-based models later.

In the first section we present a brief overview about stochastic processes in general and Markov processes in particular, especially those based on additive processes. We then proceed with defining agent-based models in terms of Markov processes and highlight in an example how the mathematical formalism can help us defining a model with the properties we want. Finally, we study established simulation techniques for stochastic processes and how they can be utilized for the realization of agent-based models.

## 1.1 Markov Processes

In this section we will introduce the notation for well-known concepts to describe and formalize the stochastic processes that will be used for the models in this thesis. Due to the length of this thesis, we will focus here only on the basic definitions and properties. For a detailed overview we refer to [19–23].

Our focus will be the class of Markov processes, which are the stochastic processes that can be described as memoryless. The memorylessness property means that the future development of the process only depends on the present state and is stochastically independent of the past development of the process conditioned on the present. This property is also known as Markov property and thus eponymous for the class of processes. In the literature there often is the distinction between discrete-time Markov chains and

continuous-time Markov processes as well as discrete and continuous state spaces. We will use a general framework that includes the different cases in one notation [13]. In the first subsection we will state the most important properties of Markov processes and introduce the notion of the transfer operator semigroup and the infinitesimal generator as a way to characterize and define Markov processes.

In the following subsections we will focus on the different cases of continuous, discrete and hybrid state spaces and illustrate how the characteristics of the randomness are varying in each case and how the specific processes are connected to the general framework of the first subsection.

### 1.1.1 General Concepts

In the following we will assume that the reader is familiar with the basic concepts of probability theory, functional analysis and stochastic analysis. We will begin with the basic definition of the stochastic process as a family of random variables.

**Definition 1.** A *stochastic process* is a family of random variables  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  defined on a common probability space  $(\Omega, \mathcal{E}, \mathbb{P})$  and mapping onto the measurable space  $\mathbb{X} \subseteq \mathbb{R}^n$  equipped with the Borel- $\sigma$ -Algebra  $\mathcal{B}(\mathbb{X})$ .  $\mathbb{X}$  is called the *state space*. If the index set  $\mathbb{T} = \mathbb{N}$ , then  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  is called a process in *discrete time*. For  $\mathbb{T} = [0, \infty)$  we call  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  a process in *continuous time*.

We will consider in principle the more general case of continuous-time processes and the case of discrete time will be mostly considered when constructing discretizations of the continuous objects. For the analysis of continuous-time processes it is reasonable to have the number of discontinuities in the sample paths restricted such that we do not have uncountably many jumps. This can be guaranteed by the càdlàg property for sample paths, that ensures that we can almost surely consider the set of discontinuities to be countable.

**Definition 2.** Let  $\mathbb{X}$  be a metric space and  $\mathbb{T} \subseteq \mathbb{R}$ . A function  $f : \mathbb{T} \rightarrow \mathbb{X}$  is called a *càdlàg function* if it is right-continuous with left limits.  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  has *càdlàg paths* if the function  $\mathbf{X}(t)$  is a càdlàg-function almost surely.

The abbreviation càdlàg is derived from the literal translation of the two conditions that need to be satisfied into the french language: **continué à droite limité à gauche**.

The next concept that is important for the analysis of stochastic processes is concerning how we encode the information about the development of a process. Figuratively speaking, we want to be able to ask questions like "How did the process evolve until time  $t$ ?". In a mathematical sense this means that we want to be able to formulate conditional probabilities and expectations with respect to the  $\sigma$ -algebras generated by the past development of a process.

**Definition 3.** Let  $(\Omega, \mathcal{E}, \mathbb{P})$  be a probability space,  $I$  a totally ordered index set and  $\mathcal{F}_i$  be a sub- $\sigma$ -algebra of  $\mathcal{E}$  for every  $i \in I$ . Then,  $(\mathcal{F}_i)_{i \in I}$  is called a *filtration* if  $\mathcal{F}_j \subseteq \mathcal{F}_k$  for all  $j \leq k$ .

Often it is necessary to consider a filtration that is adapted to a specific stochastic process. This can be guaranteed by considering the filtration that is generated by the process.

**Definition 4.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a stochastic process and  $\mathcal{F}_t$  the sub- $\sigma$ -algebra generated by the development of the process until time  $t$ , i.e.

$$\mathcal{F}_t := \sigma(\{X_s \mid s \leq t\}).$$

Then, the family  $(\mathcal{F}_t)_{t \in \mathbb{T}}$  is the filtration generated by the process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ . If  $(\mathcal{F}_t)_{t \in \mathbb{T}}$  is completed by the null-sets of  $(\Omega, \mathcal{E}, \mathbb{P})$ , then we call it the *standard filtration* of  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ .

The addition of the null sets to the filtration generated by a stochastic process is important to ensure measurability even for arbitrary initial distributions [23]. In practice this will be hardly relevant but for the theoretical foundation it is necessary to take the null sets into account.

Let us now introduce two important classes of stochastic processes, where the key properties that characterize them are defined in terms of the information of the development of the process that is encoded in the filtration. The first kind of process we consider is the martingale process which has a major application area in the modeling of fair games, which is also the origin of the name martingale itself.

**Definition 5.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a stochastic process and  $(\mathcal{F}(t))_{t \in \mathbb{T}}$  be the standard filtration of the process.  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  is called a *martingale* if  $\mathbb{E}(|\mathbf{X}(t)|) < \infty$  for all  $t \in \mathbb{T}$  and it holds that

$$\mathbb{E}(\mathbf{X}(t+s) \mid \mathcal{F}_t) = \mathbf{X}(t)$$

almost surely for all  $t, s \in \mathbb{T}$ .

The martingale property is a statement about the expected future development of the process conditioned on the past and important for the proofs of many convergence results, e.g. the dominated convergence theorem or  $L^p$ -convergence [20]. Especially in the construction of the Itô integral the martingale property is essential. For stochastic integration also a generalization of martingales is of interest.

**Definition 6.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a stochastic process and  $(\mathcal{F}(t))_{t \in \mathbb{T}}$  the standard filtration of the process. Then,  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  is called a *semimartingale* if it can be decomposed into a process  $(\mathbf{M}(t))_{t \in \mathbb{T}}$  that is a martingale and a process  $(\mathbf{A}(t))_{t \in \mathbb{T}}$  of bounded variation.

Semimartingales are closed under linear combinations. For all processes  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  that are semimartingales a stochastic integral  $\int f(t)d\mathbf{X}(t)$  for Lipschitz continuous functions  $f$  can be constructed [20]. The process defined by the stochastic integral is itself a semimartingale with càdlàg paths [20].

Now we focus on another class of processes that will be the main component of our later models. The defining property is that of randomness without memory.

**Definition 7.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a stochastic process with standard filtration  $(\mathcal{F}(t))_{t \in \mathbb{T}}$ . The process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  fulfills the Markov property if

$$\mathbb{P}(\mathbf{X}(t+s) \in A \mid \mathcal{F}(t)) = \mathbb{P}(\mathbf{X}(t+s) \in A \mid \mathbf{X}(t))$$

almost surely for all  $A \in \mathcal{B}(\mathbb{X})$  and  $s, t \in \mathbb{T}$ .

The Markov property gives us a characterization of memorylessness in the sense that the future development of the process does not depend on the whole past encoded in the filtration but only on the current state. To show that a process fulfills the Markov property can be difficult, so we also consider an equivalent characterization via the transition function that is somewhat more tangible.

**Definition 8.** A function  $p : \mathbb{T} \times \mathbb{X} \times \mathcal{B}(\mathbb{X}) \rightarrow [0, 1]$  is called *Markov transition function* if it has the following properties:

1.  $x \mapsto p(t, x, A)$  is measurable for all  $t \in \mathbb{T}$  and  $A \in \mathcal{B}(\mathbb{X})$ ;
2.  $A \mapsto p(t, x, A)$  is a probability measure for all  $t \in \mathbb{T}$  and  $x \in \mathbb{X}$ ;
3.  $p(0, x, \mathbb{X} \setminus \{x\}) = 0$  for all  $x \in \mathbb{X}$ ;
4.  $p$  fulfills the Chapman-Kolmogorov equation for all  $t, s \in \mathbb{T}, x \in \mathbb{X}$  and  $A \subseteq \mathbb{X}$ :

$$p(t+s, x, A) = \int_{\mathbb{X}} p(t, x, dz) p(s, z, A) dx.$$

We call a function  $\mathcal{Q} : \mathbb{T} \times \mathbb{X} \times \mathcal{B}(\mathbb{X}) \rightarrow [0, 1]$  that fulfills properties 1 and 2 but not necessarily 3 and 4 a *Markov kernel* [23].

We can define a stochastic process by a Markov transition function and it is especially the fourth property of the Chapman-Kolmogorov equation that is closely connected to the Markov property.

**Definition 9.** A Markov process is a stochastic process on state space  $\mathbb{X}$ , that is defined by a Markov transition function  $p : \mathbb{T} \times \mathbb{X} \times \mathcal{B}(\mathbb{X}) \rightarrow [0, 1]$  through

$$p(t, x, A) = \mathbb{P}(X_{t+s} \in A \mid X_s = x),$$

for all  $t, s \in \mathbb{T}, x \in \mathbb{X}$  and  $A \subseteq \mathbb{X}$ . If  $p(t, x, A)$  does not depend on  $t$  the process is called *homogeneous*, else it is called *inhomogeneous*.

It can be shown that a stochastic process that is defined by a Markov transition function fulfills the Markov property. Actually, it can even be shown that this relation is 1-1 [22] and thus the transition function indeed gives an equivalent characterization of Markov processes. Unfortunately, for most processes the transition function can not be written down explicitly.

Even though there are very prominent examples of stochastic processes that fulfill both the martingale and the Markov property in general both properties are independent from each other, i.e., there exist processes that fulfill only one but not the other of the two properties. However, there is a property that is closely related to both and gives us a class of stochastic processes that have both properties.

**Definition 10.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a stochastic process. If for any  $n \in \mathbb{N}$  the random variables  $\mathbf{X}(t_1) - \mathbf{X}(t_0), \mathbf{X}(t_2) - \mathbf{X}(t_1), \dots, \mathbf{X}(t_n) - \mathbf{X}(t_{n-1})$  are independent for all  $t_0 < t_1 < \dots < t_n \in \mathbb{T}$ , then  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  has the property of *independent increments*.

From the property of independent increments it follows that also the Markov property holds. On the other hand, if  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  is a process with independent increments and  $\mathbb{E}(\mathbf{X}(t))$  exists for all  $t \in \mathbb{T}$ , then the process  $(\mathbf{Y}(t))_{t \in \mathbb{T}}$  with  $\mathbf{Y}(t) := \mathbf{X}(t) - \mathbb{E}(\mathbf{X}(t))$  fulfills the martingale property [20]. So, it follows that a process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  with independent increments and  $\mathbb{E}(\mathbf{X}(t)) = 0$  for all  $t \in \mathbb{T}$  is a Markov process as well as a martingale. The other direction, however, is not true as there exist processes without independent increments that still are martingales and possess the Markov property as well.

**Definition 11.** We call a stochastic process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  with  $\mathbf{X}(0) = 0$  a.s. independent increments and càdlàg paths an *additive process*. If the increments of  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  are also stationary, i.e.,  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  is time-homogeneous, then it is called a *Lévy process*.

As the name suggests it holds that the sum of two independent additive processes is again an additive process. Also the sum of two independent Lévy processes is again a Lévy process.

Now let us recall some important properties of Markov processes that also highlight why Markov processes are interesting for modeling in general. The Markov property itself of course leads to advantages in simulation effort as independence of the past leads to less memory allocation requirements and also implies under some additional conditions convergence of the process to an equilibrium with respect to the transition function.

**Definition 12.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a Markov process. Then, the probability measure  $\mu$  is called *invariant measure* for  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  if for all  $A \subset \mathbb{X}$  holds:

$$\int_{\mathbb{X}} p(t, x, A) d\mu(x) = \mu(A).$$

If  $\mu$  can be expressed in terms of a probability density, we will also refer to  $\mu$  as invariant density.

Often we are not interested in a single trajectory of a Markov process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  but rather in the development of an ensemble or a probability density on the state space. For the propagation of a density we use the concept of the transfer operator [13].

**Definition 13.** Let  $p : \mathbb{T} \times \mathbb{X} \times \mathcal{B}(\mathbb{X}) \rightarrow [0, 1]$  be the transition function of a Markov process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  and  $\mu$  an invariant measure for  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ . Then, the mapping  $P^t : L^r(\mu) \rightarrow L^r(\mu)$  for  $t \in \mathbb{T}$  and  $1 \leq r < \infty$  is defined as

$$\int_A P^t \nu(y) d\mu(y) = \int_{\mathbb{X}} \nu(x) p(t, x, A) d\mu(x).$$

$P^t$  is called *propagator* or forward transfer operator.

**Proposition 1.** The set of propagators  $P^t : L^r(\mu) \rightarrow L^r(\mu)$  of a Markov process with transition function  $p : \mathbb{T} \times \mathbb{X} \times \mathcal{B}(\mathbb{X}) \rightarrow [0, 1]$  is a semigroup.

*Proof.* First it is to show, that  $P^0$  is the identity on  $L^r(\mu)$ .

$$\int_A P^0 \nu(y) d\mu(y) = \int_{\mathbb{X}} \nu(x) p(0, x, A) d\mu(x) = \int_{\mathbb{X}} \nu(x) \chi_{A \cap \{x\}} d\mu(x) = \int_A \nu(x) d\mu(x)$$

Thus,  $P^0 \nu = \nu$  and therefore  $P^0 = Id_{L^r(\mu)}$ .

The second property we need to check is  $P^{t+s} = P^t P^s$  for all  $s, t \in \mathbb{R}^+$ . This follows from the Chapman-Kolmogorov equation, which is fulfilled by the transition function.

$$\begin{aligned} \int_A P^{t+s} \nu(y) d\mu(y) &= \int_{\mathbb{X}} \nu(x) p(t+s, x, A) d\mu(x) \\ &= \int_{\mathbb{X}} \int_{\mathbb{X}} \nu(x) p(s, x, dz) p(t, z, A) d\mu(x) \\ &= \int_{\mathbb{X}} P^s \nu(z) p(t, z, A) \mu(dz) \\ &= \int_A P^t P^s \nu(y) d\mu(y). \end{aligned}$$

□

**Definition 14.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a continuous-time Markov process with transfer operator semigroup  $P^t$  and  $\mathcal{C}^0(\mathbb{X})$  the space of continuous functions on  $\mathbb{X}$ . If

1. for all continuous functions  $\nu \in \mathcal{C}^0(\mathbb{X})$  also  $P^t \nu \in \mathcal{C}^0(\mathbb{X})$
2. for all  $\nu \in \mathcal{C}^0(\mathbb{X})$  and all  $x \in \mathbb{X}$  the limit  $\lim_{t \rightarrow 0} P^t \nu = \nu$  exists,

then  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  is called a *Feller process* and  $P^t$  a *Feller semigroup*.

All Feller processes  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  have a representation with càdlàg paths and Feller semigroups are strongly continuous [23]. For strongly continuous semigroups we can define an infinitesimal generator of the semigroup. The generator can be seen as a generalization of the derivative for operator semigroups as it describes the change of a density in an infinitely small time interval.

**Definition 15.** Let  $P^t : L^r(\mu) \rightarrow L^r(\mu)$  be the propagator of a continuous-time Markov process with  $1 \leq r \leq \infty$  and let  $D(L)$  the set of all  $\nu \in L^r(\mu)$  such that the limit

$$L\nu = \lim_{t \rightarrow 0} \frac{P^t\nu - \nu}{t}$$

exists. The operator  $L : D(L) \rightarrow L^r(\mu)$  defined in this way is called *infinitesimal generator* of the semigroup  $P^t$ .

Additive and Lévy processes possess a strongly continuous transfer operator semigroup as they have càdlàg paths and thus can be characterized by a generator.

For the propagator semigroup of a Markov process we have  $P^t\mu = \mu$  for all  $t$  and for the generator  $L\mu = 0$  holds. However, not every Markov process possesses an invariant measure. There are some restrictions on either the state space  $\mathbb{X}$  or the dynamics of  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ . We will recite two important results on that matter. The first result requires the notion of tightness of a sequence of measures, which tells us that the probability mass is concentrated on compact sets.

**Definition 16.** A sequence  $M$  of probability measures on a topological space  $\mathbb{X}$  is called *tight* if for every  $\varepsilon > 0$  there exists a compact set  $K \subset \mathbb{X}$  such that  $\nu(K) \geq 1 - \varepsilon$  for every  $\nu \in M$ .

With this definition in mind we can state the result from Krylov and Bogolyubov, which states a condition under which a Feller process has an invariant measure [24].

**Theorem 17.** (*Krylov-Bogolyubov*) Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a Feller process on a polish space  $\mathbb{X}$  with propagator semigroup  $P^t$ . If there exists an element  $x \in \mathbb{X}$  such that the sequence of measures  $(P^t\delta_x)_{t \geq 0}$  is tight, then there exists an invariant probability measure  $\mu$  for  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ .

A consequence of the theorem is that Feller processes on compact spaces  $\mathbb{X}$  have an invariant probability measure  $\mu$  [24]. This already is applicable in many modeling scenarios and especially in the agent-based model case the space in which agents can move will be chosen compact. For the case of spaces  $\mathbb{X}$  we can apply another criteria that is based on Lyapunov functions.

**Definition 18.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a Markov process on a polish space  $\mathbb{X}$  with transition function  $p$ . A measurable function  $V : \mathbb{X} \rightarrow \mathbb{R}_+ \cup \{\infty\}$  is called a *Lyapunov function* for  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  if it satisfies the following conditions:

- there exist  $x \in \mathbb{X}$  for which  $V(x)$  is finite, i.e.,  $V^{-1}(\mathbb{R}_+) \neq \emptyset$
- the level sets  $V^{-1}(\{x \leq c\})$  are compact for every  $c \in \mathbb{R}_+$
- there exists a positive constant  $\gamma < 1$  and a constant  $C \in \mathbb{R}$  such that

$$\int_{\mathbb{X}} V(y)p(t, x, y)dy \leq \gamma V(x) + C$$

for all  $x \in \mathbb{X}$  with  $V(x) < \infty$  and all  $t \in \mathbb{T}$ .

The second condition implies that  $V$  is asymptotically converging towards infinity outside of compact sets and the third condition ensures that the dynamics of the process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  are non-explosive. For Feller processes  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  that admit a Lyapunov function there exists an invariant measure. [24] There are also more general Lyapunov criteria for processes that are not Feller [22].

For Markov processes that have an invariant measure we also have a convergence of the process in a probabilistic sense.

**Definition 19.** A Markov process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  with invariant measure  $\mu$  is called *ergodic* with respect to  $\mu$  if for all functions  $f : \mathbb{X} \mapsto \mathbb{R}$  with  $\int_{\mathbb{X}} f(x) d\mu(x) < \infty$  it holds that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t f(\mathbf{X}(t)) dt = \int_{\mathbb{X}} f(x) d\mu(x)$$

almost surely.

If a Markov process has a unique invariant measure, then it also is ergodic [24]. Ergodicity then implies that the time average of an infinitely long trajectory of a process converges to the ensemble average of an infinite ensemble of realizations, which is distributed according to the invariant measure  $\mu$ . In this sense a Markov process converges to the invariant measure  $\mu$  if it exists and is unique.

It is often of interest when a Markov process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  enters some area of the state space for the first time, e.g., for modeling the spreading of infections when is the first time an infected individual arrives at a new location. This can be mathematically formulated as a stopping time of the process.

**Definition 20.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a Markov process and  $A \subset \mathbb{X}$ . Then, the first hitting time of the set  $A$  when starting in  $x$  is defined as:

$$\tau_A(x) = \inf\{t \geq 0 | X_t \in A, X_0 = x\}$$

The expectation of the first hitting time of  $A$  when starting in  $x$  is denoted by:

$$m_A(x) = \mathbb{E}(\tau_A(x)).$$

$m_A$  is called *mean first passage time* or mean first hitting time.

Analogous to the arrival of the process at a set, it is also often of interest when the process leaves a set of interest. To stay with the example of infection spreading the first time an infectious person leaves an area would be an exit event of interest.

**Definition 21.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a Markov process and  $A \subset \mathbb{X}$  open and connected. Then, for all  $x \in A$  the first exit time from  $A$  is defined as

$$\rho_A(x) = \inf \left\{ t \geq 0 : \int_0^t \chi_{A^c}(X_s) ds > 0, X_0 = x \right\}.$$

We define the exit rate  $\lambda_A := \frac{1}{\rho_A(x)}$

If  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  has continuous trajectories, this is equivalent to the first hitting time of the complement of  $A$ .

Another interesting quantity that is based on first hitting times is the committor function [13].

**Definition 22.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a Markov process and  $A, B \subset \mathbb{X}$ . Then, the committor function of  $A$  and  $B$  is defined as

$$q_{AB}(x) := \mathbb{P}(\tau_B(x) < \tau_A(x)),$$

which is the probability that  $B$  is hit by  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  before  $A$  when starting in  $x$ .

The value of this function describes whether the point  $x$  is closer to the set  $A$  or the set  $B$  with respect to the dynamics of the process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ . The committor function can be used to characterize rare transitions of a process and to identify transition pathways [13]. Thus, it will be an important tool when we address in a later section the topic of metastability and model reduction.

The concepts that we have discussed so far are rather abstract as they apply to many different kinds of Markov processes. We will now continue with more details on specific processes where we can also give some intuitive examples for the general concepts of this section.

### 1.1.2 Markov Diffusion Processes

The first kind of randomness without memory that we consider is the Brownian Motion in a continuous state space  $\mathbb{X}$ . In this subsection all processes are in continuous time. The fundamental characteristic of the Brownian motion is a change of direction at every point in time, independent of the previous direction. The following characterization of (standard) Brownian motion  $(\mathbf{B}(t))_{t \in \mathbb{T}}$  is the most common:

1.  $\mathbf{B}(0) = 0$  almost surely.
2.  $\mathbf{B}(t)$  is almost surely continuous.
3.  $\mathbf{B}(t)$  has independent increments.
4.  $\mathbf{B}(t)$  has stationary increments that follow a normal distribution, i.e.,  $\mathbf{B}(t+s) - \mathbf{B}(s) \sim \mathcal{N}(0, t)$  for  $t, s \geq 0$ .

As  $(\mathbf{B}(t))_{t \in \mathbb{T}}$  has independent increments and  $\mathbb{E}(\mathbf{B}(t)) = \mathbb{E}(\mathbf{B}(t) - \mathbf{B}(0)) = 0$  we have that  $(\mathbf{B}(t))_{t \in \mathbb{T}}$  is a Markov process as well as a martingale. There exists a representation of the Brownian motion such that the paths of the Brownian motion are almost surely continuous but also almost surely nowhere differentiable [23]. The Brownian motion thus fulfills also the conditions of a Lévy process. For the Brownian motion we can even write down the transition function explicitly.

**Example 1.** Let  $(\mathbf{B}(t))_{t \in \mathbb{T}}$  be a Brownian motion in  $\mathbb{R}$ . Then, the increments  $\mathbf{B}(t) - \mathbf{B}(s)$  are independent and distributed according to  $\mathcal{N}(0, t-s)$  for all  $s < t \in \mathbb{T}$ . Therefore

$$\begin{aligned} \mathbb{P}(\mathbf{B}(t+s) \in A | \mathbf{B}(s) = x) &= \mathbb{P}(\mathbf{B}(t+s) - \mathbf{B}(s) \in \{y-x | y \in A\}) \\ &= \int_A \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{|y-x|^2}{2t}\right) dy. \end{aligned}$$

This means the transition function of the Brownian motion is given by

$$p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{|y-x|^2}{2t}\right).$$

The standard Brownian motion in  $\mathbb{R}$  can be easily extended to  $n$  dimensions by defining an independent one-dimensional Brownian motion for each dimension. The increments are then distributed according to an  $n$ -dimensional Gaussian distribution. In the figures 1.1 and 1.2 we illustrate the Brownian motion for one and two dimensions.

Next we will consider stochastic processes  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  that are composed of a deterministic part and a scaled Brownian motion as added randomness. Usually, these processes are defined by a stochastic differential equation, which is in fact rather an integral equation with a stochastic integral term but written down in differential notation [25].

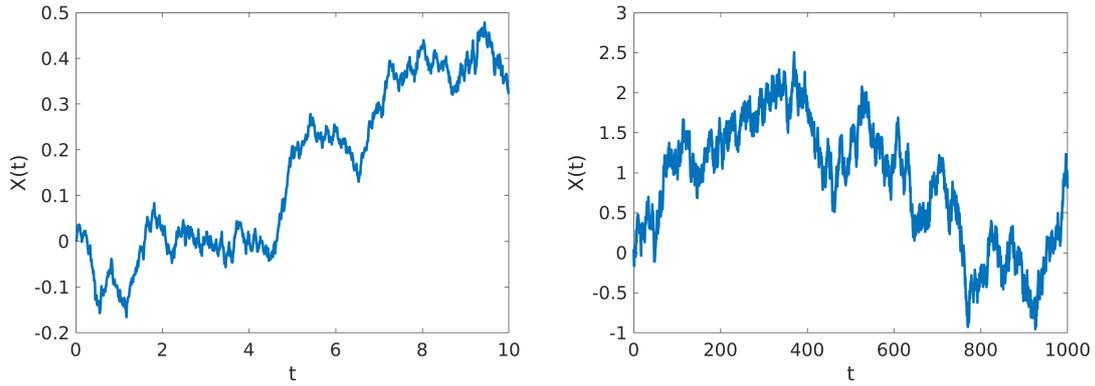


Figure 1.1: Standard Brownian motion in one dimension for a short (left) and a long trajectory (right). The characteristic change of direction at (almost) every point in time is clearly visible, also the increasing variance with time. An infinitely long trajectory of Brownian motion will cross any finite threshold almost surely.

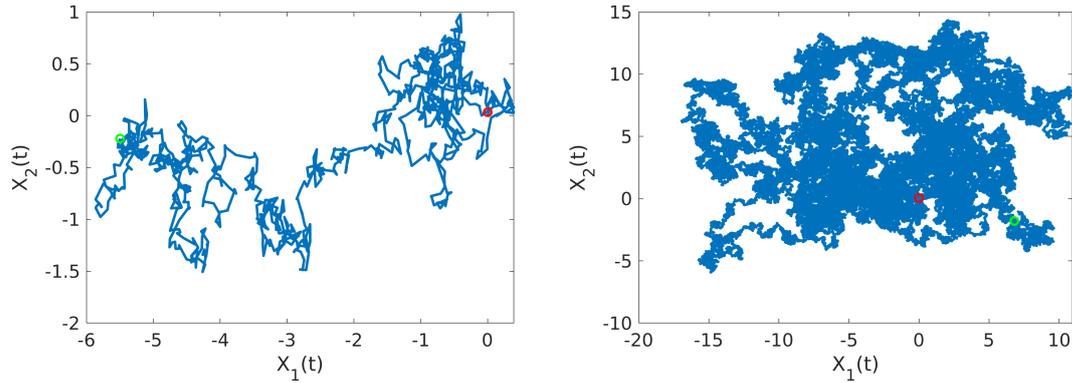


Figure 1.2: Standard Brownian motion in two dimensions for a short (left) and a long (right) trajectory with starting point chosen to be the origin marked in red and end the point of the trajectory in green. Similarly to the 1-dimensional case an infinite trajectory would leave any ball with finite radius around the origin almost surely but also return to a neighborhood of the origin countably many times. In higher dimensions however the Brownian motion is transient.

**Definition 23.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a stochastic process with state space  $\mathbb{X} \subseteq \mathbb{R}^d$  satisfying the stochastic differential equation

$$d\mathbf{X}(t) = b(\mathbf{X}(t))dt + \sigma(\mathbf{X}(t))d\mathbf{B}(t)$$

with  $b, \sigma$  being Lipschitz continuous functions and  $(\mathbf{B}(t))_{t \in \mathbb{T}}$  a Brownian motion in  $\mathbb{R}^d$ .  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  is called *Markov diffusion process* with drift  $b$  and diffusion coefficient  $\sigma$ .

The easiest example of a diffusion process is the case where both  $b$  and  $\sigma$  are just constant functions. The resulting diffusion process is called Brownian motion with drift and can be seen as a linear function with added noise. Even for a comparably small drift constant one can observe that the deterministic part dominates the long term evolution of the process (compare Figure 1.3).

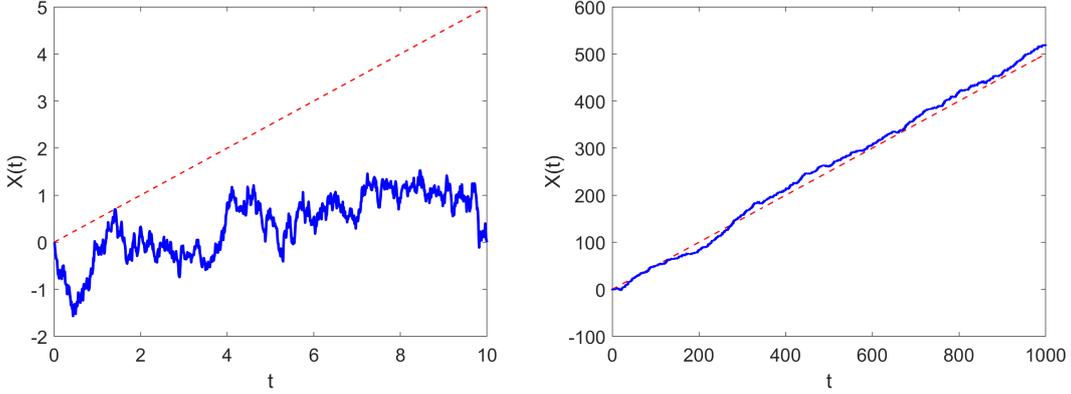


Figure 1.3: Visualization of a short (left) and long (right) trajectory of a one-dimensional Brownian motion with drift for  $b = 0.5$  and  $\sigma = 1$ . For the short term development the added randomness of the Brownian motion is dominating as we have chosen the diffusion parameter  $\sigma$  to be higher as the drift parameter  $b$ . In the long term development the deterministic drift term (marked with a red dashed line) is leading.

Neither the standard Brownian motion ( $b = 0, \sigma = 1$ ) or a Brownian motion with drift possess a stationary density as we have  $\mathbf{X}(t) \sim \mathcal{N}(bt, \sigma^2 t)$  changing with the advancement of the time  $t$ . Even though the standard Brownian motion does not possess a stationary probability density the set  $B_\epsilon(0)$  that denotes the ball with radius  $\epsilon > 0$  is a recurrent set (in the case of 2 dimensions or less) [26], highlighting that recurrence alone does in general not imply the existence of a stationary probability density.

A dynamically more interesting case is the diffusion in a potential landscape given by a smooth function  $V$  (see Figure 1.4 for an example), which is utilized in many models for particle physics, e.g., Langevin or Smoluchowski models [13].

**Example 2.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a Markov diffusion process, where the dynamics are defined by the stochastic differential equation

$$d\mathbf{X}(t) = -\nabla V(\mathbf{X}(t))dt + \sigma d\mathbf{B}(t)$$

with  $\sigma \in \mathbb{R}$ ,  $\mathbf{B}(t)$  a Brownian motion in  $\mathbb{R}^n$  and  $V$  a smooth function. Then, the development of the density  $\partial_t \nu_t$  of  $\mathbf{X}(t)$  is given by the Fokker-Planck equation

$$\partial_t \nu_t = L \nu_t$$

for all  $t \in \mathbb{T}$ . The generator  $L$  of  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  is therefore given by

$$L = -\nabla V(x)\nabla_x + \frac{\sigma^2}{2}\Delta_x,$$

where  $\nabla_x$  is the gradient and  $\Delta_x$  the Laplacian operator with respect to  $x$ .

In the case of  $\nabla V = 0$  and  $\sigma = 1$  the stochastic differential equation (SDE) describes the standard Brownian motion in  $\mathbb{R}^n$  and the corresponding generator is  $\frac{1}{2}\Delta_x$ .

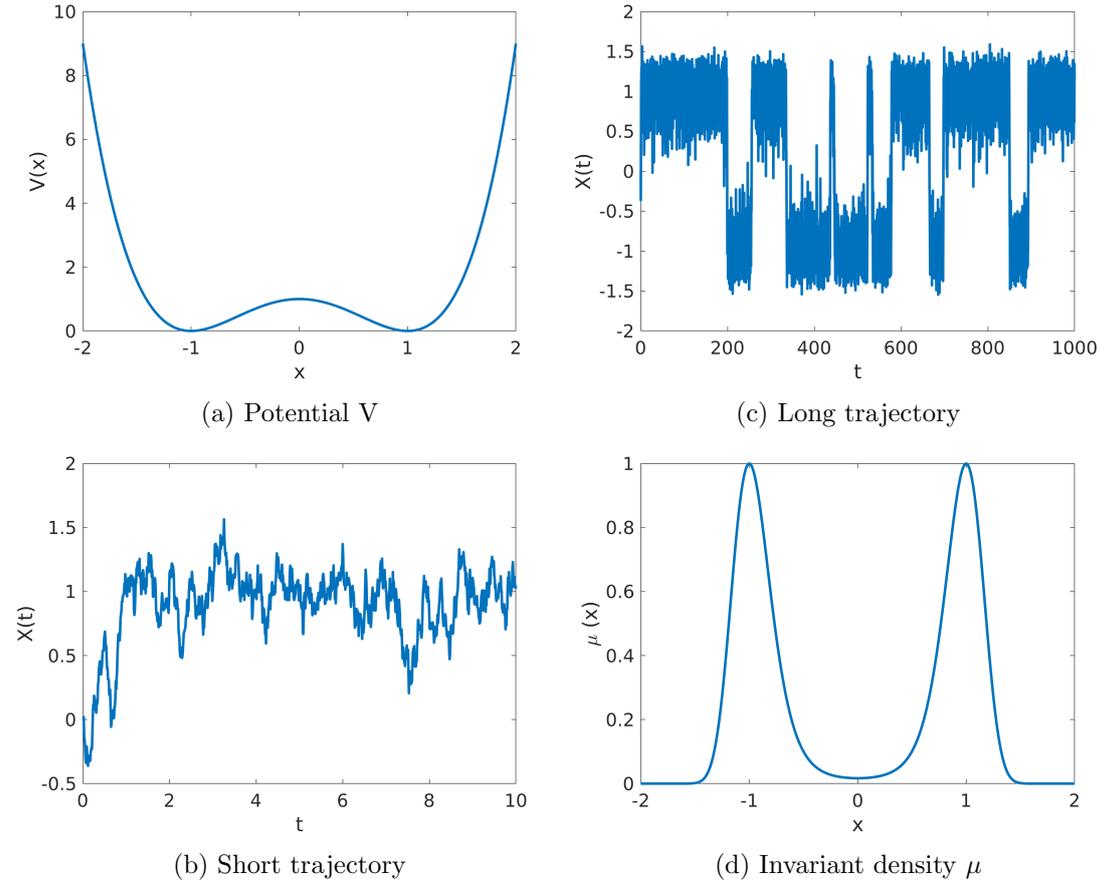


Figure 1.4: Diffusion process in a double-well potential  $V$  which is shown in (a). The process starting in  $\mathbf{X}(0) = 0$  quickly moves towards one of the two local minima of  $V$  and stays in a local neighborhood for a longer time which illustrated in (b) for a short trajectory of a numerical realization. In the long trajectory of the same realization (c) we can observe that transitions between the two wells are possible but rather rare events and that the process is ergodic with respect to the invariant density  $\mu$  given by the Boltzmann distribution that is depicted in (d).

For a diffusion process in a potential landscape with a constant diffusion coefficient the invariant probability density function exists if  $V$  is a Lyapunov function and is then given by the Boltzmann distribution [13]

$$\mu(x) = \frac{1}{Z} \exp\left(-\frac{2}{\sigma^2} V(x)\right)$$

with  $Z = \int_{\mathbb{X}} \exp\left(-\frac{2}{\sigma^2} V(x)\right) dx$  as a scaling constant.

### 1.1.3 Markov Jump Processes

Next we want to focus on continuous-time Markov processes with random events only occurring at discrete time points and not continuously. While in the continuous case the memorylessness of the diffusion processes is in some sense inherited from the Brownian motion, in the discrete case the memorylessness will be characterized by exponentially distributed waiting times between state changes of the process. We call a state change after a waiting time a *jump event* and the corresponding Markov process a *jump process*. We will first focus on countable state spaces that allow us to write down the generator of the process in terms of a (countable) matrix.

**Definition 24.** Let  $\mathbb{S}$  be a countable set. A (countable) matrix  $L = (\lambda_{ij})_{i,j \in \mathbb{S}}$  with entries that fulfill

$$\begin{cases} \lambda_{ij} \geq 0 \text{ for } i, j \in \mathbb{S} \text{ with } i \neq j \\ \sum_{j \in \mathbb{S}} \lambda_{ij} = 0 \end{cases}$$

is called a *transition rate matrix*.

With the notion of a rate matrix we can define Markov jump processes as the class of Markov processes that are characterized by transition rate matrices.

**Definition 25.** A continuous-time Markov process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  with countable state space  $\mathbb{S}$  is called Markov jump process if the infinitesimal generator  $L$  has the form of a transition rate matrix. The entries  $\lambda_{ij}$  are the transition rates determining the transition probabilities from state  $i$  to state  $j$  for  $i \neq j$  and the entries  $-\lambda_{ii}$  are the exit rates from state  $i$ .

The construction of the process from the rate matrix is straightforward. The exit rate  $-\lambda_{ii}$  determines the holding time  $\tau \sim \text{Exp}(-\lambda_{ii})$  describing how long the process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  stays in a discrete state  $i \in \mathbb{S}$ . After this exponential waiting time a jump event occurs and the state of the process is changed from state  $i$  to another state  $j$  according to the probability determined by the transition rate  $\lambda_{ij}$ , i.e.,  $\mathbb{P}(\mathbf{X}(\tau) = j) = \frac{\lambda_{ij}}{|\lambda_{ii}|}$ . If the entries of the rate matrix depend on the time  $t$  we call the corresponding jump process inhomogeneous in contrast to the homogeneous case, where the jump and transition rates are constant.

A well-known class of Markov jump processes are counting processes, often denoted by  $(\mathbf{N}(t))_{t \in \mathbb{T}}$ , that take values in  $\mathbb{N}$  and have increments of 1 [21]. The state  $\mathbf{N}(t)$  of the process at time  $t$  can be interpreted as the number of jump events that have happened until time  $t$ . These processes can also be characterized as Poisson point processes on the real line [21] and the standard Poisson process with unit rate is very useful as a basic process for the construction of more complicated jump processes. Our definition of the standard Poisson process will be in terms of our framework as a Markov jump process with a generator matrix.

**Definition 26.** The Markov jump process  $(\mathcal{P}(t))_{t \in \mathbb{T}}$  with state space  $\mathbb{N}$  defined by the countable rate matrix with entries

$$\lambda_{ij} = \begin{cases} -1 & \text{if } i = j \\ 1 & \text{if } j = i + 1 \\ 0 & \text{else} \end{cases}$$

with  $\mathcal{P}(0) = 0$  is called *unit rate Poisson process*.

From the standard Poisson process we can construct any counting process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  by adapting the rate matrix according to an intensity rate function  $\lambda$  that is different from the unit rate. We can accomplish this also if we utilize the unit rate Poisson process and rescale the time  $t$  according to the intensity  $\lambda$  instead [27]. In general, we distinguish between the case of a constant intensity rate  $\lambda > 0$  and the case of a time-dependent intensity rate given by a positive and measurable function  $\lambda(t)$ .

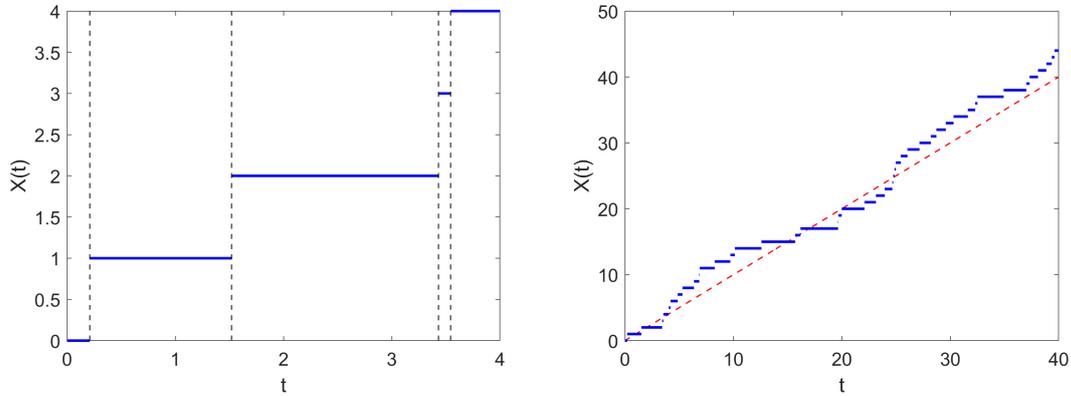


Figure 1.5: Visualization of a short (left) and longer (right) trajectory of a unit rate Poisson process. The jump times are marked with dashed black lines in the short trajectory and the graph of the identity on  $\mathbb{R}$ , which is the expectation of the process, is plotted with a dashed red line in the longer trajectory.

**Definition 27.** Let  $\lambda > 0$  and  $(\mathcal{P}(t))_{t \in \mathbb{T}}$  a unit rate Poisson process. The counting process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  defined by  $\mathbf{N}(t) := \mathcal{P}(\lambda t)$  is called a *homogeneous Poisson process* with *intensity rate*  $\lambda$ .

The homogeneous Poisson process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  is illustrated in Figure 1.5 and has the following key properties, which also give an alternative definition of the process:

1.  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  is a Lévy process, i.e., it has stationary and independent increments and càdlàg paths.
2. The random variable  $\mathbf{N}(t)$  is Poisson distributed with parameter  $\lambda t$ .

Now we consider the case of a time-dependent intensity rate function.

**Definition 28.** Let  $\lambda : \mathbb{T} \rightarrow [0, \infty)$  be a locally integrable function such that  $\int_{t_1}^{t_2} \lambda(\tau) d\tau < \infty$  holds for all  $t_1, t_2 > 0$  and  $(\mathcal{P}(t))_{t \in \mathbb{T}}$  be a unit rate Poisson process. The counting process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  defined by  $\mathbf{N}(t) := \mathcal{P}(\int_0^t \lambda(s) ds)$  is called an *inhomogeneous Poisson process* with *intensity rate function*  $\lambda$ .

The inhomogeneous Poisson process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  with time-dependent rate function  $\lambda$  has the following key properties:

1.  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  is an additive process with càdlàg paths and independent but not stationary increments.
2. The random variable  $\mathbf{N}(t)$  is Poisson distributed with parameter  $\int_0^t \lambda(s) ds$ .

Similarly, we can also construct state-dependent inhomogeneous Poisson processes with rate functions defined on  $\mathbb{N}$  or  $\mathbb{N} \times \mathbb{T}$ . The key properties are analogous to the case of a rate function defined on  $\mathbb{T}$ .

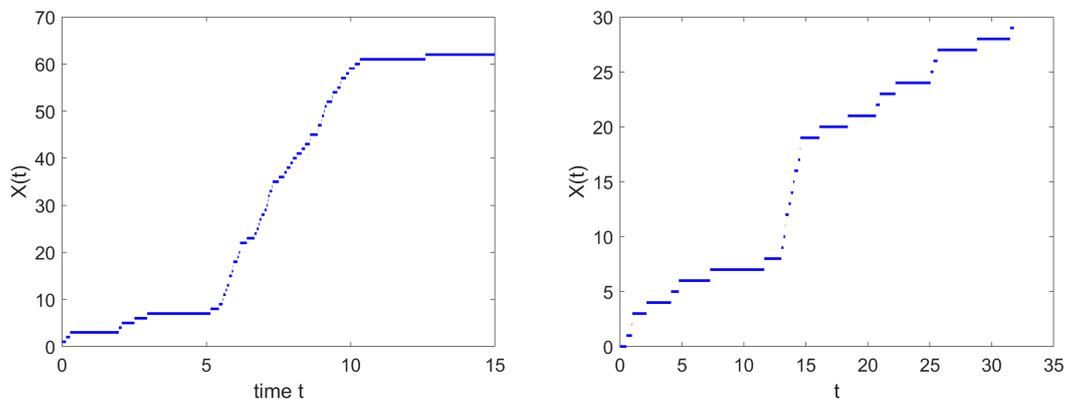


Figure 1.6: Visualization of a time-dependent (left) and state dependent (right) inhomogeneous Poisson process. In the time-dependent example the intensity switches between 1 and 10 every 5 units of time, in the state-dependent example the intensity switches between 1 and 10 every 10 events.

Since Poisson processes are semimartingales, we can construct a stochastic integral with respect to the unit rate Poisson process and introduce an SDE representation for counting processes [28]. For a counting process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  with intensity rate function  $\lambda$  we can write

$$\mathbf{N}(t) = \mathbf{N}(0) + \int_0^t \lambda(s) d\mathcal{P}(s)$$

or

$$d\mathbf{N}(t) = \lambda(t) d\mathcal{P}(t)$$

with  $(\mathcal{P}(t))_{t \in \mathbb{T}}$  being a unit rate Poisson process.

Through the composition of Poisson processes we can construct any Markov jump process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  with finite state space  $\mathbb{S}$  with  $n$  elements given by a rate matrix  $L$  by defining for each possible transition from  $i$  to  $j$  a state change vector  $v_{ij}$  and an independent inhomogeneous Poisson process  $\mathcal{P}_{ij} \left( \int_0^t \delta_i(\mathbf{X}(\tau)) \lambda_{ij}(s) ds \right)$ . It then holds that

$$\mathbf{X}(t) = \mathbf{X}(0) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n v_{ij} \mathcal{P}_{ij} \left( \int_0^t \delta_i(\mathbf{X}(\tau)) \lambda_{ij}(s) ds \right), \quad (1.1)$$

where  $\delta_i$  denotes the indicator function of the discrete state  $i \in \mathbb{S}$ . If we restrict the number of state change vectors that define the possible jump events to be finite (e.g., in the case of a counting process we have only increments by 1 and thus only one state change vector), then we can also write down a jump process with a countable state space  $\mathbb{S}$  in a similar way.

For now we have only considered a finite set of state change vectors for possible jump events. The next generalization includes the possibility for random instead of deterministic state change vectors. This also enables the construction of jump processes that take values in a continuous state space beyond a discrete subset.

**Definition 29.** Let  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  be a counting process with intensity  $\lambda$  and  $\mathcal{Q}$  a Markov kernel on  $\mathbb{X} \times \mathcal{B}(\mathbb{X})$ . A *compound Poisson process*  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  is defined by

$$\mathbf{X}(t) := \sum_{i=0}^{\mathbf{N}(t)} Z_i$$

with  $Z_0 = \mathbf{X}(0)$  and  $Z_i$  distributed according to  $\mathcal{Q}(\mathbf{X}(\tau_i))$  for  $i > 0$  and  $\tau_i$  denoting the times of the jump events.

We can also denote the process given by (1.1) as a compound Poisson process by defining a Markov kernel  $\mathcal{Q}$  with

$$\mathcal{Q}(i, v_{ij}) = \frac{\lambda_{ij}}{|\lambda_{ii}|}$$

and  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  with a state dependent intensity rate function

$$\lambda(\mathbf{X}(t)) := \sum_{\substack{j=1 \\ j \neq \mathbf{X}(t)}}^n \lambda_{\mathbf{X}(t)j}.$$

For compound Poisson processes there is also a notation utilizing a stochastic integral with respect to  $(\mathbf{N}(t))_{t \in \mathbb{T}}$ . We can write

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t Z_{\mathbf{N}(s)} d\mathbf{N}(s).$$

We now proceed with a generalization of Markov jump processes to piecewise-deterministic Markov processes, which are driven by Poisson noise but also include a deterministic influence.

### 1.1.4 Piecewise-Deterministic Markov Processes

For Markov jump processes the state of the process does not change between two jump events. We now want to consider stochastic processes, where the randomness is generated via a Poisson process, but where the state of the system also can change between jumps in a deterministic way. These processes are called piecewise-deterministic Markov processes (PDMP) and the term has first been introduced by Davis in [29]. His construction of a PDMP consists of three main concepts:

1. A deterministic flow  $\Phi$  that governs the development of the process between two jump events.
2. An inhomogeneous Poisson process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  that has an intensity rate function  $\lambda$  for the times of jump events.
3. A Markov kernel  $\mathcal{Q}$  that determines the probability of different jump events, depending on the current state of the process at the time of a jump event.

With the concepts from the previous section we can define the process as a compound Poisson process with an additional deterministic influence.

**Definition 30.** Let  $\mathbb{X} \subset \mathbb{R}^d$ ,  $\mathbb{S}$  a discrete set and  $\mathbb{Y} = \mathbb{X} \times \mathbb{S}$ . Let  $\lambda$  be a measurable function from  $\mathbb{Y}$  to  $[0, \infty)$ ,  $\mathcal{Q}$  a Markov kernel on  $\mathbb{Y} \times \mathcal{B}(\mathbb{Y})$  and  $\Phi(x, i, t)$  a differentiable function on  $\mathbb{X} \times \mathbb{T}$  for every  $i \in \mathbb{S}$ . The process  $(\mathbf{Y}(t))_{t \in \mathbb{T}}$  given by

$$\mathbf{Y}(t) = \Phi(\mathbf{Y}(t), t) + \sum_{j=0}^{\mathbf{N}(t)} Z_j$$

with  $Z_i$  distributed according to  $\mathcal{Q}(\mathbf{Y}(t))$  and  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  being a Poisson process with intensity  $\lambda(\mathbf{Y}(t))$  is called *piecewise-deterministic Markov process*.

We have expanded the continuous state space  $\mathbb{X}$  with a discrete index set, such that we can include switches in the deterministic flow after a jump event leads to a change in the index set. For each element of the index set we have a distinct branch, i.e., a copy of  $\mathbb{X}$ . A trajectory of the process takes only values in one branch at a time, so we can interpret the trajectory as a motion in  $\mathbb{X}$  with the index set being necessary to encode flow changes but only in the background. For the development of a probability distribution through the transfer operator or generator of the dynamics we are considering an object that is a combination of a probability density for  $\mathbb{X}$  in the first component and a probability vector for the index set  $\mathbb{S}$  in the second component. The derivation of a generator for a PDMP is also discussed in the original paper of Davis [29].

We will consider now a PDMP in a slightly simpler setting presented in [30] that will be sufficient for our modeling purposes and can also be expressed with a time change representation. We restrict the number of different possible jump events to be finite and also the domain for jumps to the discrete component of the state space  $\mathbb{S}$  and thus we

can write down a representation similar to (1.1). We again consider state change vectors  $v_k$  for each of the possible changes at jump events and split the intensity rate function  $\lambda$  into a total of  $J$  different intensity rate functions  $\lambda_k$ , one for each state change vector. For this setting we can define  $(\mathbf{Y}(t))_{t \in \mathbb{T}}$  as

$$\mathbf{Y}(t) = \Phi(\mathbf{Y}(t), t) + \sum_{j=1}^J v_j \mathcal{P}_j \left( \int_0^t \lambda_j(\mathbf{Y}(s)) ds \right)$$

with  $\mathcal{P}_j$  being independent unit rate Poisson processes. For such a process we can also formulate a stochastic differential equation with a deterministic part and a part that is integrated with respect to Poisson processes.

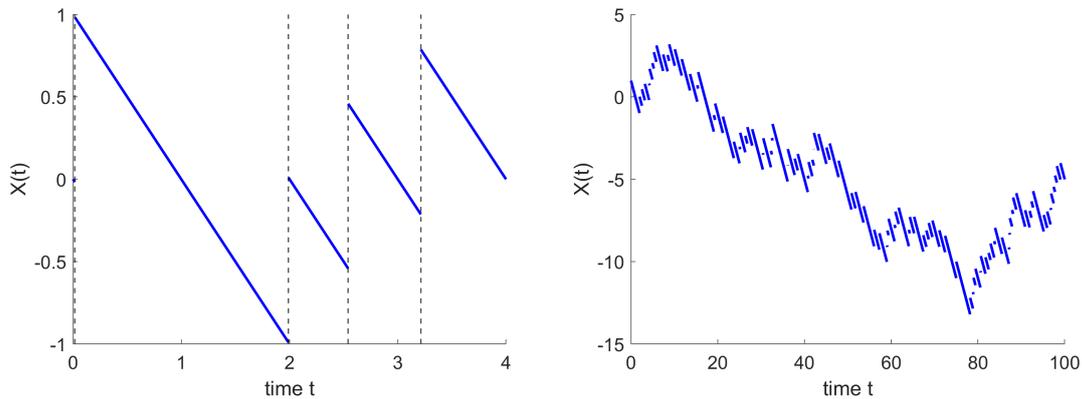


Figure 1.7: A short (left) and long (right) trajectory from a numerical realization of a unit rate compensated Poisson process.

**Example 3.** Let  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  be a Poisson process with intensity rate function  $\lambda$ . Then, the stochastic process defined by

$$\tilde{\mathbf{N}}(t) := - \int_0^t \lambda(s) ds + \mathbf{N}(t) \tag{1.2}$$

is called *compensated Poisson process*. The integral term  $\int_0^t \lambda(s) ds$  of this process is also called *compensator* and is the expected number of jumps until time  $t$ . As we have

$$\mathbb{E}(\tilde{\mathbf{N}}) = \mathbb{E}(\mathbf{N}(t)) - \int_0^t \lambda(s) ds = \int_0^t \lambda(s) ds - \int_0^t \lambda(s) ds = 0$$

for all  $t \in \mathbb{T}$  in addition to the property of independent increments and càdlàg paths the compensated Poisson process is a martingale. Because of the martingale property the compensated Poisson process plays an important role in the theory of stochastic integration as it can be used to prove a version of Itô's lemma for stochastic integrals with respect to Poisson processes. [20]

The homogeneous compensated Poisson process is closely related to the Brownian motion even though the characteristics of the paths are fundamentally different. Both processes are martingales and the variance of the distribution of the process at time  $t$  grows linearly for both cases. For sufficiently large values of  $t$  the Poisson distribution with parameter  $\lambda t$  can be approximated by a normal distribution, which is then in the case of a compensated Poisson process again centered at 0 due to the compensation term. So, the limiting distribution of a homogeneous compensated Poisson process is the same as that of the Brownian motion. This result is known as the Poisson approximation and can also be utilized to approximate Poisson noise with a high intensity through the randomness of a Brownian motion [31].

**Example 4.** Let  $\mathbb{S} := \{1, 2\}$ ,  $\mathbb{X} := \mathbb{R}$  and  $\mathbb{Y} := \mathbb{X} \times \mathbb{S}$ . We define  $(\mathbf{Y}(t))_{t \in \mathbb{T}}$  on  $\mathbb{Y}$  to be the PDMP given by

$$\mathbf{Y}(t) = (\mathbf{X}(t), \mathbf{S}(t)) = \mathbf{Y}(0) + \int -1^{\mathbf{S}(t)} dt + \sum_{j=1}^J v_j \mathcal{P}_j \left( \int_0^t \lambda_j(\mathbf{Y}(s)) ds \right) \quad (1.3)$$

with  $v_1 = (0, 1)$ ,  $v_2 = (0, -1)$  being the  $J = 2$  state change vectors and

$$\lambda_i(\mathbf{Y}(t)) = \begin{cases} 1 & \text{if } S(t) \neq i \\ 0 & \text{else} \end{cases}$$

for  $i = 1, 2$  the corresponding intensity rate functions. The process switches between the 2 discrete states after an exponential waiting time and is either linearly increasing or decreasing between two switch events. A numerical realization of the process is visualized in Figure 1.8.

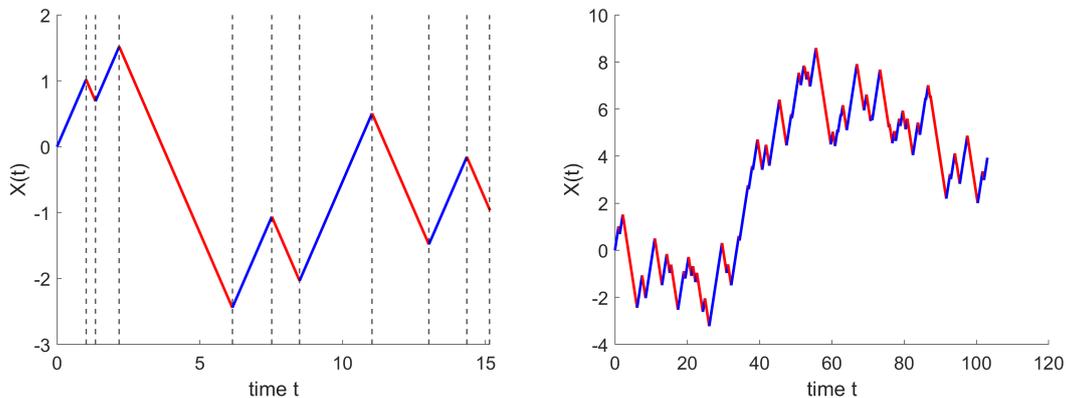


Figure 1.8: Piecewise linear PDMP example for a short (left) and long (right) trajectory. While  $S(t) = 1$  the line is colored blue and while  $S(t) = 2$  it is colored red.

### 1.1.5 Jump Diffusion Processes

In the last sections we have considered stochastic differential equations that include a deterministic drift term either in combination with Poisson noise in the piecewise-deterministic case or with a Brownian noise term in the diffusion case. The next logical step is to construct processes that include both types of noise. This could be interpreted either as starting with a PDMP and substituting the deterministic development between jump events with a diffusive development and thus the process becoming piecewise-diffusive or as starting with a diffusion process and adding discontinuous jumps to the continuous dynamics and thus the process becoming a jump diffusion. Since the term jump diffusion process is much more common we will also use it to describe these kind of dynamics. Common application areas for jump diffusion processes are mathematical finance or molecular dynamics (e.g., stock/option pricing or reaction diffusion models) and of course the application that we focus on in this thesis: agent-based models.

**Definition 31.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a stochastic process that is given by a stochastic differential equation

$$d\mathbf{X}(t) = b(\mathbf{X}(t), t)dt + \sigma(\mathbf{X}(t), t)d\mathbf{B}(t) + Z_{\mathbf{N}(t)}d\mathbf{N}(t) \quad (1.4)$$

with  $b, \sigma$  Lipschitz continuous functions,  $(\mathbf{B}(t))_{t \in \mathbb{T}}$  a Brownian motion that is independent from the inhomogeneous Poisson process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  and  $Z_{\mathbf{N}(t)}$  distributed according to a Markov kernel  $\mathcal{Q}$  for all times  $t$ . We call  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  a *jump diffusion process*.

Most of the time we will, similarly as in the PDMP case, make the simplification of a finite set of state change vectors determining the possible jump events and can thus use the somewhat simpler notation with transition vectors and their corresponding Poisson processes instead of a compound Poisson process.

**Example 5.** As a relatively simple example we consider an expansion of the piecewise-linear process from example 4 by adding a standard Brownian motion term and a compound Poisson term with increments distributed according to a standard normal distribution. We can then write down the process as

$$\mathbf{Y}(t) = \mathbf{Y}(0) + \left( \int -1^{\mathbf{S}(t)} dt, 0 \right) + (\mathbf{B}(t), 0) + \sum_{k=0}^{\mathcal{R}(t)} Z_k + \sum_{j=1}^J v_j \mathcal{P}_j \left( \int_0^t \lambda_j(\mathbf{Y}(s)) ds \right)$$

with  $\mathcal{R}$  being a unit rate Poisson process independent from  $\mathbf{B}$  and  $\mathcal{P}_j$  and  $Z_k$  normally distributed random variables. A sample path of the process is visualized in Figure 1.9.

An interesting result concerning jump diffusion processes is the Lévy-Itô decomposition [28], which essentially states that every Lévy process can be decomposed into three parts:

1. a linear deterministic drift part
2. a scaled Brownian motion part

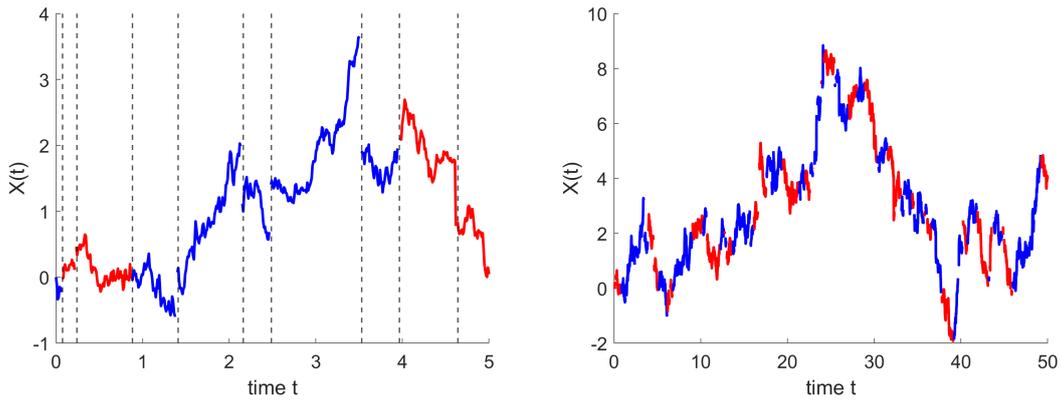


Figure 1.9: Visualisation of a short (left) and long (right) trajectory from a numerical realization of the jump diffusion process in example 5. For the short trajectory dotted black lines are used to indicate the jump times of the compound Poisson jumps in  $\mathbb{X}$  as well as the discrete switches in  $\mathbb{S}$ . While  $S(t) = 1$  the deterministic drift term is positive and the line is colored blue, while  $S(t) = 2$  the drift term is negative and the line is colored red.

3. a jump part consisting of a compensated and a compound Poisson process.

For the rigorous statement see Theorem 42 in [20]. A similar result can also be proven for additive processes [28, 32], which can also be decomposed into a deterministic, a diffusion and a jump part. In this case the drift and diffusion coefficients do not need to be constant but still state independent. So, the class of jump diffusion processes actually is sufficient to describe all additive stochastic processes as those are characterized exactly by these three building blocks. Theorem 32 in the book of Protter [20] states a condition for processes defined by stochastic differential equations to possess the Markov property. It is sufficient that all integrands of the SDE are additive processes and all coefficients are Lipschitz continuous functions.

We now proceed with the introduction of agent-based models and a mathematical formalization of them in terms of Markov processes.

## 1.2 Mathematical Formalization of Agent-Based Models

Mathematical modeling is a possibility for comprehending the reality that we live in. A complicated real-world object or process can be represented by an abstraction that we can formulate in the mathematical language based on observations. The study of this somewhat simpler object can help us understand some aspects of the real-world counterpart for which it is a model. Of course even with the best possible model we will always be limited by our human perception of reality and human logic. This insight ultimately leads to the famous quote that is attributed to George Box [33]: "all models are wrong, but some are useful". As mathematicians or scientists in general we are of course interested in formulating useful models which need to carefully balance simplicity and realism. A model that is very realistic can be too complex for any meaningful analysis, e.g., if there are too many parameters to assess, which of them are the parameters responsible for certain results. On the other hand, a model that makes too many simplifications might not describe the problem accurately enough. In general, it is a good strategy to start with a model that is more on the simplistic end of this spectrum and expand it step by step to add more realism but in a manner such that too much complexity is avoided and it is clear which effect on the system the additional element causes. This is also one reason to consider Markov processes as a possible restriction for the building blocks of a model as an extension to a model with another Markov process does not change the fundamental mathematical properties.

Agent-based models are on this spectrum rather on the more realistic side but they can easily be constructed as too complex for a meaningful analysis. Restricting oneself to ABMs that can be written down in a formal mathematical way is a first step to avoid too complex models but even then one often ends up with high-dimensional stochastic processes that have no analytical solution, especially when real-world data is involved. So, usually the strength of ABMs lies in hypothesis testing through Monte Carlo simulations and a subsequent computational analysis. As this can lead to a high computational effort, it is crucial to construct models in a way such that they can be efficiently simulated or else not many scenarios or parameter settings can be explored. This is another reason to consider a mathematical formalism within the realm of Markov processes as this enables us to utilize efficient simulation algorithms and established model reduction techniques.

In this section we will discuss how ABMs are defined in other works and provide a formalism for ABMs that can be written as jump diffusion processes. Then, we will introduce an example that can be seen as a basic building block for interacting agent systems that will later also be used to illustrate how model reduction can be applied to ABMs. In another example we utilize the mathematical formalism to define an ABM such that a desired equilibrium emerges only from self organization of the agents according to local information.

### 1.2.1 Definition of Agent-Based Models

The first question that needs to be answered in this section is rather obvious, the answer to it though requires some elaboration. The question is of course: what is an agent-based model? Definitions of ABMs range from very descriptive formulations [34] to very precise formulations using concepts from models in statistical physics [35]. Which level of mathematical formalism is used heavily depends on the application area and the mathematical background that is expected of the reader. Often an interdisciplinary audience is the target as ABMs can be viewed as a bridge between disciplines [36]. The range of application areas is vast but the most prominent are the disciplines concerned with the interactions of human individuals within societies, e.g., social science [4], economics [5], epidemiology [6], ecology [3] or archaeology [7]. However, ABMs are not limited to human interactions as they can also be utilized to model the decision-making and behavior of animals, e.g., in flocking dynamics [34] or movements for foraging [37], and the behavior of particles in molecular dynamics [38]. What is connecting these different applications is the interaction of the entities that are modeled with each other and with an environment in which they are existing. Interaction alone is, however, not unique to ABMs as they share this trait with other model types like compartmental ODE and metapopulation models that can be used for modeling population dynamics, e.g., for disease spreading within a society [39,40]. The difference between these different model types for this example is the modeling scale which ultimately leads to different dynamical descriptions for the same real-world process. While compartmental ODE models provide often a macroscopic description of a process, ABMs are focusing on the microscopic dynamics. The microscopic description of an interaction process features at least some discreteness in the sense that the distinction between the interacting entities is always possible. So, a first descriptive answer to the opening question would be: An agent-based model is a microscale model for the interactions of discrete entities with each other and with their environment.

Now let us be a bit more specific and first take a closer look at the properties a discrete entity should have to be considered an agent. The state of an agent in general consists of multiple variables that each represent a certain feature relevant to the application, e.g., a position in space or an indicator about an opinion or the presence/absence of specific information or physical goods. So, to be precise an agent is a collection of state variables that can take values in continuous as well as discrete spaces [35]. A system of agents is consequently the collection of variables of all agents. The change of these state variables is how we define the interactions of agents with each other and their environment and thus an agent-based model is essentially a dynamical system. In principle we could write down any function that acts on the space on which the variables defining the agent system are defined and to specify the rules after which the agents behave. However, to get a meaningful model that can be analyzed it is reasonable to restrict oneself, e.g., by having the function defining the rules not depending on the individual agent itself but only on the state of the system. This raises then the question: if agents do not behave according to individual rules but rather general rules, are they then still distinguishable from each

other? There are at least two general cases in which this is still true, the first one being the case where the ABM is essentially a network model with each agent represented by a node on a network with a non-trivial network topology. The second case would be those of at least one of the defining variables of an agent being continuous, such that all agents will have a different value almost surely. We will be interested mostly in the second case in this work. On another note using general behavior rules does not exclude heterogeneity among the agents, as different agent types can be encoded by a discrete state variable. So, in this work we will only use the term agent-based model for (random) dynamical systems with a state space that is defined by a collection of variables that contains at least one continuous variable and usually also at least one discrete variable to allow for the definition of heterogeneous behavior.

Though a position can also be defined on a discrete grid or network and an opinion could also be defined to take values in a continuous opinion space [41], we will later without too much loss of generality assume that all variables concerning the *mobility* of an agent (e.g., position, velocity, orientation) are continuous while variables describing the *status* (e.g., susceptible, infected, indicator of knowledge/opinion) of an agent are discrete. So, in this interpretation the microscale characteristic of an ABM refers to the spatial resolution of the model. An ABM for continuous opinion dynamics is then to be viewed as movement in an opinion space. The environment in which the agents move and interact can be defined implicitly, either through shared state variables of agents or the functions determining the behavior and interaction rules, and does not need to be an additional entity itself. For network models with agents that are not mobile the network topology can be defined through the interaction functions as well, making network models for social interaction essentially a special case of the more general case of an ABM with mobile agents.

As we now have a good picture of what we mean by the term agent-based model in this work, we can continue to focus on more specific types of ABMs. In particular we are interested in ABMs that can be described by Markov processes. The inclusion of randomness in ABMs is very common as in the case of the complex real-world systems that we try to model we are always dealing with missing information, especially when it comes to human decision-making. The class of Markov processes is on the one hand broad enough to include many interesting ABMs based on established physical models that utilize Markov processes [35], on the other hand the restriction to processes that are memoryless has huge analytical and computational advantages. Whether the assumption of memoryless randomness is justified will be in the end more a question of the existence of a better alternative assumption. Following the principle of Ockham's razor, I believe that Markovian randomness is in general the lesser assumption than the alternative as we would then have to make an additional assumption about the memory characteristics of the random process, which themselves would require some additional justification. Including memory in the decision-making of agents can also be realized by expanding the state space and leaving the fundamental characteristic of randomness Markovian. Especially when in the process of building a model, we want to start simple

and expand the model iteratively through extensions using models based on additive Markov processes is a reasonable decision.

We now proceed with the formalization of ABMs that can be described by Markov processes and introduce the notation for ABMs that will be used in all following examples. A similar formalization has been done in [16] and we will follow the overall structure and notation.

## 1.2.2 Formalization of ABMs as Markov Processes

### Definition of the agent system

The state space for defining the collection of variables representing a single agent will be the product space of a compact domain  $\mathbb{X} \subset \mathbb{R}^d$  for encoding continuous variables and a discrete set  $\mathbb{S} \subseteq \mathbb{N}$  for encoding all discrete variables. Usually, we consider a finite number of  $n_s$  different possibilities for the status of an agent. The combined state space will be denoted by  $\mathbb{Y} := \mathbb{X} \times \mathbb{S}$ . Without loss of generality we refer to the continuous variables of an agent  $\alpha$  as the *position*  $x_\alpha$  and the change of the continuous variables as *mobility* of an agent. We call the collection of discrete variables of an agent  $\alpha$  the *status*  $s_\alpha$  and refer to status changes also as *adoption* of a new status. Both the mobility and the adoption dynamics will be defined as Markov processes in continuous time, i.e., the index set  $\mathbb{T}$  is a continuous time interval, e.g.,  $\mathbb{T} = [0, \infty)$ .

We consider a finite set of agents with the number of agents  $n_a$  being constant. In a later application we will also introduce variable agent numbers but with an upper bound for the maximum number of agents. Through an extra status variable for agents that encodes whether an agent is currently existing or not the case of variable agent numbers can be reduced to the case of constant agent numbers as long as the number of agents is bounded.

The whole system of the  $n_a$  agents is encoded through the two vectors collecting the variables of all agents: the vector  $X = (x_\alpha)_{\alpha=1, \dots, n_a}$  consisting of all positions and the vector  $S = (s_\alpha)_{\alpha=1, \dots, n_a}$  denoting the status of all agents. We call the vector  $Y = (X, S)$  the *system state* and define the space of all possible system states as  $\mathbb{Y} := \mathbb{X}^{n_a} \times \mathbb{S}^{n_a}$ .

### Change of state variables

The discrete status changes of an agent will be modeled by a Markov jump process with finitely many different possible jump events referred to as *adoption/status change events*. As human mobility can exhibit Levy walk characteristics [42] as well as characteristics of Brownian motion [43] we assume the mobility process to be in general a jump diffusion process. For the application of model reduction by projection of the dynamics to a Markov state model we will restrict ourselves later to a pure Markov diffusion process for the mobility process and omit thereby the possibility for compound Poisson jumps of the continuous state variable. We thus have a strict separation between discrete dynamics in the discrete space  $\mathbb{S}$  and continuous dynamics in the continuous space  $\mathbb{X}$ . These can nevertheless still be coupled and affect each other. However, under certain assumptions for the jump characteristics there also exist asymptotic convergence results and characterizations of metastable behavior for jump diffusions in potential landscapes [44], so under the right assumptions our model reduction framework from the next chapter can also be applied in the more general case.

## Mobility process

We formulate the jump diffusion process for the agent-based model in terms of the system state and define  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  as the (jump) diffusion process on  $\mathbb{X}^{n_a}$  that governs the movements of all  $n_a$  agents. The motion of a single agent is then related to the respective marginal process of  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ . As we have chosen  $\mathbb{X}$  to be compact we will have an invariant measure for the distribution of the agents' positions as long as the dynamics of  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  are those of a Feller process.

The environment in which the agents move is implicitly defined in the stochastic differential equation generating the mobility process and not a separate entity. External changes to the environment through time (e.g., climate changes) can be modeled in terms of the movements of the agents depending on the time variable in addition to the positions and status of all agents. An example for this would be a diffusion process in a time changing potential landscape. In that case, however, we have in general non-equilibrium dynamics.

## Adoption process

For the adoption process of a single agent  $\alpha$  we consider a finite number of  $J$  possible adoption events, each associated with a *state change vector*  $v_j \in \mathbb{Y}$  and an *adoption rate function*

$$f_j^{(\alpha)} : \mathbb{Y} \rightarrow [0, \infty)$$

specifying the intensity of the jump process. In the case of  $\mathbb{S}$  being a finite set, i.e.,  $s_\alpha \in \mathbb{S} := \{1, \dots, n_s\}$ , we can write down a jump process for each possible transition of the form  $i \rightarrow j$  for  $i, j \in \mathbb{S}$  with associated adoption rate function

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

and with associated state change vector

$$v_{ij}^{(\alpha)} := (0_{\mathbb{X}^{n_a}}, (j - i)e_\alpha) \quad (1.5)$$

where  $0_{\mathbb{X}^{n_a}}$  denotes the 0-vector in  $\mathbb{X}^{n_a}$  and  $e_\alpha$  denotes the  $n_a$ -dimensional vector with the  $\alpha$ th entry being 1 and all other entries being 0, i.e., the  $\alpha$ th unit vector of  $\mathbb{R}^{n_a}$ . For a given system state  $Y = (X, S)$  we have

$$Y + v_{ij}^{(\alpha)} = (X, S - ie_\alpha + je_\alpha)$$

so only the status variable of agent  $\alpha$  gets assigned a new value.

While in general the status changes of an agent can depend on the whole system state  $Y$ , we often define the adoption rate functions in a way that not all the position and status values of other agents are influencing the intensity of the jump process. The general case can be interpreted as having global knowledge or interacting on a global scale, while a reduction of dependencies can be interpreted as having locality for both information about and interaction with other agents. We now introduce two important types of adoption rate functions, which are the core building blocks of many models.

### First order adoptions

The first type is for modeling status changes of an agent that are independent of the state of all other agents. The adoption rate function  $f_{ij}^{(\alpha)}$  then only depends on the position and status of the agent  $\alpha$  itself. We call this type of status changes *first order adoptions* and can write

$$f_{ij}^{(\alpha)}(X, S) = \delta_i(s_\alpha) \gamma_{ij}(x_\alpha), \quad (1.6)$$

where  $\delta_i$  denotes the indicator function of a status, i.e.,  $\delta_i(s) = 1$  if  $s = i$  and  $\delta_i(s) = 0$  otherwise, and  $\gamma_{ij} : \mathbb{X} \rightarrow [0, \infty)$  specifies the intensity rate for a status transition from  $i$  to  $j$  depending on the spatial location of the acting agent. In particular, we set  $\gamma_{ii}(x) = 0$  for all positions  $x$ .

### Second order adoptions

The next type of adoption dynamics we consider is based on pairwise interactions with other agents. In this case we speak of *second order adoptions* and we set

$$f_{ij}^{(\alpha)}(X, S) = \delta_i(s_\alpha) \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{n_\alpha} \delta_j(s_\beta) \gamma_{ij}(x_\alpha, x_\beta), \quad (1.7)$$

where  $\gamma_{ij} : \mathbb{X}^2 \rightarrow [0, \infty)$  is a function defining the intensity rate for a status adoption from  $i$  to  $j$  depending on the positions of two interacting agents. In a more special setting, this rate  $\gamma_{ij}(x_\alpha, x_\beta)$  depends only on the distance between the interacting agents, e.g., they need to be closer than some *interaction radius*  $r > 0$  to interact, as for example in the Doi model [45] in the context of chemical reaction systems. The underlying idea is that interactions of agents that can lead to status changes require proximity of the agents in physical space. For this case, we set

$$\gamma_{ij}(x_\alpha, x_\beta) := c_{ij} \cdot d_r(x_\alpha, x_\beta) \quad (1.8)$$

for a constant  $c_{ij} \geq 0$ , where  $d_r : \mathbb{X}^2 \rightarrow \{0, 1\}$  for  $r > 0$  is the distance indicator function:

$$d_r(x_\alpha, x_\beta) = \begin{cases} 1, & \text{if } |x_\alpha - x_\beta| \leq r \\ 0, & \text{otherwise.} \end{cases} \quad (1.9)$$

For  $i = j$  there are no status transitions and we set  $c_{ii} = 0$ . The way we have defined the second order adoptions the intensity of the associated jump process also depends on the number of agents with a different status in the neighborhood as we assume interactions with all neighbors. For some applications with high agent numbers in the neighborhood it can make more sense to define the adoption rate function with a threshold or to divide the adoption rate by the number of neighbors to limit the influence of the number of neighbors on the jump intensity.

## Process formulation

The coupling of the diffusion process and the jump dynamics given by status changes leads to a Markov process  $(\mathbf{Y}(t))_{t \in \mathbb{T}}$  on the system state space  $\mathbb{Y}$ . Let  $p(X, S, t)$  be the probability mass function for the process  $(\mathbf{Y}(t))_{t \in \mathbb{T}}$  to be in the system state  $(X, S)$  at time  $t$ , where the marginal with respect to  $X$  is a continuous density function and the marginal with respect to  $S$  is a probability vector.

We define for each status  $i$  an operator  $L_i$  that describes the change of the probability mass function through the motion of a single agent under the condition that it is currently in status  $i$ . Then, we can write down an operator  $L$  for the movement of the agents,

$$Lp(X, S, t) := \sum_{\alpha=1}^{n_a} L_{s_\alpha}^{(\alpha)} p(X, S, t), \quad (1.10)$$

where  $L_{s_\alpha}^{(\alpha)}$  is defined as  $L_i$  for  $i = s_\alpha$  acting on a function  $p(X)$  with respect to the component  $x_\alpha$  of  $X$  (see Example 6 for details). Note that, consequently,  $L_{s_\alpha}^{(\alpha)}$  acts only on the position part of the probability mass function  $p(X, S, t)$  in (1.10).

As for the adoption dynamics, we define

$$\begin{aligned} Gp(X, S, t) &:= - \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)}(X, S) p(X, S, t) \\ &\quad + \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)} \left( (X, S) - v_{ij}^{(\alpha)} \right) p \left( (X, S) - v_{ij}^{(\alpha)}, t \right) \\ &= - \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)}(X, S) p(X, S, t) \\ &\quad + \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)}(X, S - ie_\alpha + je_\alpha) p(X, S - ie_\alpha + je_\alpha, t), \end{aligned} \quad (1.11)$$

where  $e_\alpha$  denotes the  $\alpha$ th unit vector of  $\mathbb{R}^{n_a}$ . The first term on the right-hand side refers to the outflow from the current state through adoption events and the second term to the inflow through adoption events that would lead to the current state. The change of  $p(X, S, t)$  including movement and status transitions is then given by the set of differential equations

$$\partial_t p(X, S, t) = Lp(X, S, t) + Gp(X, S, t). \quad (1.12)$$

The strength of this general agent-based approach is that different types of restrictions regarding the interaction dynamics can be included and quite complicated dynamics can be formulated. On the other hand, the coupled differential equations which describe the system are usually not analytically solvable. Instead, Monte Carlo simulations of the

dynamics are required to sample the quantities we are interested in and the simulation of such complex systems can be numerically very costly, especially if local neighborhoods have to be computed in every time step.

We proceed with two explicit examples to illustrate how to use this framework to define agent-based models in terms of Markov processes. Afterwards, we will discuss the simulation of agent-based systems.

### 1.2.3 Guiding Example

The first example that we also call *guiding example* consists of the basic building blocks for modeling spreading processes. This example was also featured in [16]. We define for the agent mobility a diffusion process in a potential landscape that exhibits metastability and in addition a jump process for second order adoptions.

**Example 6** (Two-status dynamics in a double-well potential). We consider the continuous space for the agent movements to be  $\mathbb{X} = \mathbb{R}^2$  and the discrete status space to be  $\mathbb{S} = \{1, 2\}$ . The movement of a single agent is defined as a diffusion process and is given by the following stochastic differential equation

$$d\mathbf{x}(t) = -\left(\frac{\sigma}{2}\right)^2 \nabla U(\mathbf{x}(t))dt + \sigma d\mathbf{B}(t),$$

with  $U(x_1, x_2) = (x_1^2 - 1)^2 + 7x_2^2$  being a 2-dimensional double-well potential (see Figure 1.10), diffusion constant  $\sigma \in \mathbb{R}$  and  $\mathbf{B}(t)$  a standard Brownian motion process in  $\mathbb{R}^2$ . The movement process of all  $n_a = 100$  agents is denoted by  $\mathbf{X}(t)$ . As for the status dynamics, we consider second-order adoptions with rate functions defined by (1.7) and (1.8), assuming that only the changes from status 1 into status 2 are possible, namely with rate constant  $c_{12} > 0$ , while transitions back to status 1 are excluded by setting  $c_{21} = 0$ .

In the initial state, all agents are in status 1 except for one agent in the left well, given by the subset  $(-\infty, 0] \times \mathbb{R} \subset \mathbb{X}$ , which has status 2. The critical transition event that we are interested in is the first time that an agent with status 2 makes the transition from the left to the right well  $(0, \infty] \times \mathbb{R}$  as this enables the spreading of status 2 within the right well. In Figure 1.10b we see that this transition happens only after almost all agents in the left well have adopted status 2. This indicates that adoptions within one well happen faster than between two wells, which is due to the metastability of the diffusion process for the mobility of the agents.

In the following chapter we will focus on this type of dynamics and derive reduced models that preserve its main properties. In particular, we will consider the critical transition event in order to compare the relevant statistics and quantify the approximation error of derived reduced models. This example has also been used in [16] for the same purpose.

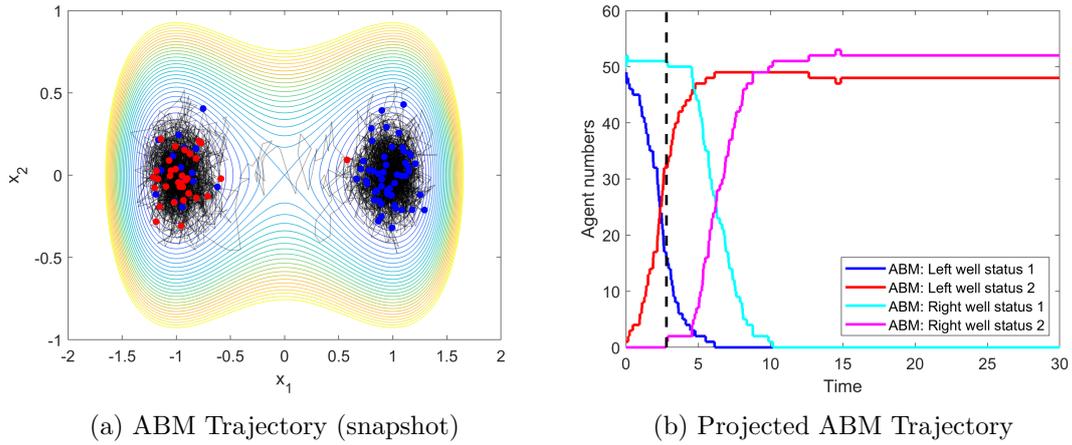


Figure 1.10: Two-status dynamics in a double-well potential, see Example 6 [16]. The adoption rate constant is chosen as  $c_{12} = 0.1$  and the interaction radius as  $r = 0.15$  (a) Snapshot of an ABM trajectory for a simulation with  $n_a = 100$  agents with contour lines of the potential  $U$  from Example 6 with diffusion constant  $\sigma = 1.2$ . The blue dots refer to spatial positions of agents in status 1, the red dots represent agents in status 2. Grey lines show movement trajectories of the agents. (b) Projected trajectory of the same example, showing the temporal evolution of the total number of agents in each of the two wells (given by  $(-\infty, 0] \times \mathbb{R}$  and  $(0, \infty) \times \mathbb{R}$ ) depending on the status. The dashed black line refers to the time of the snapshot. The example will be revisited in Section 2.2.5.

### 1.2.4 Self Organisation of Agents According to Local Information

Agent-based models often revolve around interactions based on local information [46], but some applications also require agents to act on global information about the system [47]. In this section, we will use a simple example to explore the differences between agents acting on global information and agents acting on local information. For this example we will also consider a dependence of the mobility dynamics on the status of the agents, and thus we have a feedback loop between the mobility and the adoption process [48, 49].

Let us first define the state space and the mobility process for an agent system that will serve as a numerical example in this section.

**Example 7.** We consider a system with a constant number of  $n_a = 1000$  agents, status set  $\mathbb{S} = \{1, 2\}$ , position space  $\mathbb{X} = [-2, 2]$  and  $L_i$ , for  $i = 1, 2$  as defined in Equation (1.10) given by the diffusion SDE

$$d\mathbf{x}(t) = -\nabla V_i(\mathbf{x}(t))dt + \sigma d\mathbf{B}(t)$$

with the potential  $V_i$  chosen as depicted in Figure 1.11,  $\sigma = 0.7$  and  $\mathbf{B}_t$  a one-dimensional Brownian motion. For each  $L_i$ , the invariant probability density function is given by the Boltzmann distribution

$$\mu_i(x) = \frac{1}{Z} \exp\left(-\frac{2}{\sigma^2} V_i(x)\right)$$

with  $Z = \int_{\mathbb{X}} \exp\left(-\frac{2}{\sigma^2} V_i(x)\right) dx$  as a scaling constant. The potentials are constructed so that agents with status 1 are more attracted to the left well and agents with status 2 are more attracted to the right well.

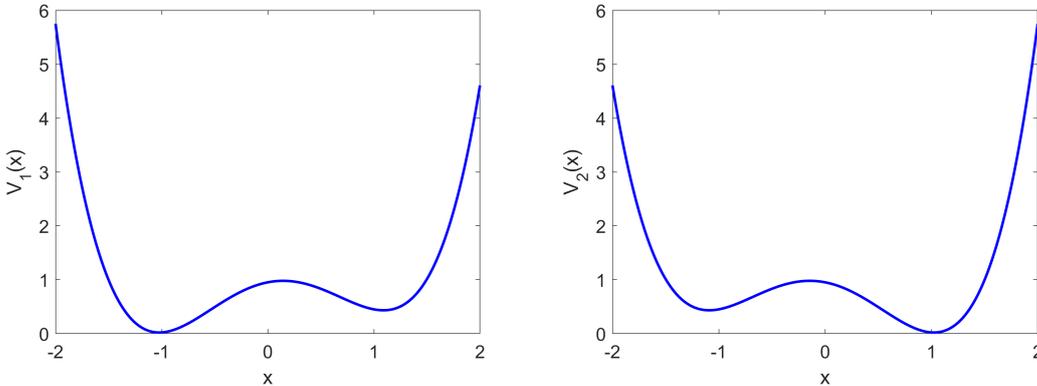


Figure 1.11: Graphs of potentials  $V_1$  (left) and  $V_2$  (right).

The basic idea for the adoption dynamics is that the agent system should organize itself such that the agents are distributed according to a specified target distribution  $\eta$ . We choose  $\eta$  to be the equilibrium distribution resulting from an ensemble of agents that

do not adapt their status and are initialized with the status distributed according to the probability vector  $\pi$ . For the numerical example, we choose  $\pi = (0.4, 0.6)$ . We distinguish between two general modeling approaches for the status dynamics. The first one considers agents with global knowledge, where we define the adoption rate functions such that they are only depending on the status of the agents and not on the positions. The second approach considers agents with local knowledge about other agents within a specified interaction radius. We can utilize the Markov process description for agent-based models to define a suitable adoption rate function for the dynamics of this case.

For the numerical example we will compare the stationary distributions of the different dynamics, sampled from 50 simulations each, since we can not compute the stationary density analytically for all cases. As a measure for the error, we will use the total variation distance<sup>1</sup> between the probability measures of the theoretical target distribution  $\eta$  and the empirical distributions obtained from the agent-based simulations.

In the following, the adoption rate functions are defined so that they are independent of time  $t$  and depend only on the system state  $Y(t)$ . We will first consider the case of agents with access to global information. Although in our example we have only two different status values, we formulate the adoption rate functions with the more general setting of a status set  $\mathbb{S}$  with  $n_s$  different status values in mind.

### $\pi$ -Global Information

Let agent  $\alpha$  be in status  $i$ ,  $\gamma > 0$  a constant for specifying the jump intensity and

$$g_{ij}^{(\alpha)}(X, S) := \gamma \max \left\{ 0, \left( \pi_j - \frac{\sum_{\beta=1}^{n_a} \delta_j(s_\beta)}{n_a} \right) \right\} \quad (1.13)$$

the adoption rate function for transitions from  $i$  to  $j$ , with  $\delta_j$  being the indicator function for status  $j$ . The adoption rate depends on the global population statistics and is either 0 or linearly increasing from 0 if the fraction of agents in status  $j$  is less than the value  $\pi_j$  of the given status distribution.

Equipped with these dynamics, the status change of the system stops completely as soon as the probability vector  $\pi$  matches the status distribution of the agent system, and thus the theoretical equilibrium distribution is given by the invariant probability density functions of the diffusion generators  $L_1, \dots, L_{n_s}$  weighted by the respective entries of  $\pi$ . We denote this stationary distribution for a single agent as

$$\begin{aligned} \eta : \mathbb{X} \times \mathbb{S} &\rightarrow [0, \infty) \\ (x, i) &\mapsto \pi_i \mu_i(x). \end{aligned}$$

---

<sup>1</sup>The total variation distance of two probability distributions  $\nu_1, \nu_2$  is defined as

$$d_{\text{TV}}(\nu_1, \nu_2) := \sup_A |\nu_1(A) - \nu_2(A)|$$

The invariant measure for the whole system is given by the product measure  $\eta^{n_a}$ . We are also interested in the local status distribution, for which we can define a function  $\pi^{(\text{loc})} : \mathbb{X} \times \mathbb{S} \rightarrow \mathbb{R}$  with

$$\pi^{(\text{loc})}(x, i) = \frac{\eta(x, i)}{\sum_{j=1}^{n_s} \eta(x, j)} \quad (1.14)$$

describing the frequency of each status  $i$  for a given location  $x$  according to the invariant distribution  $\eta$ . In the numerical example,  $\pi$  is eventually reached by construction. The

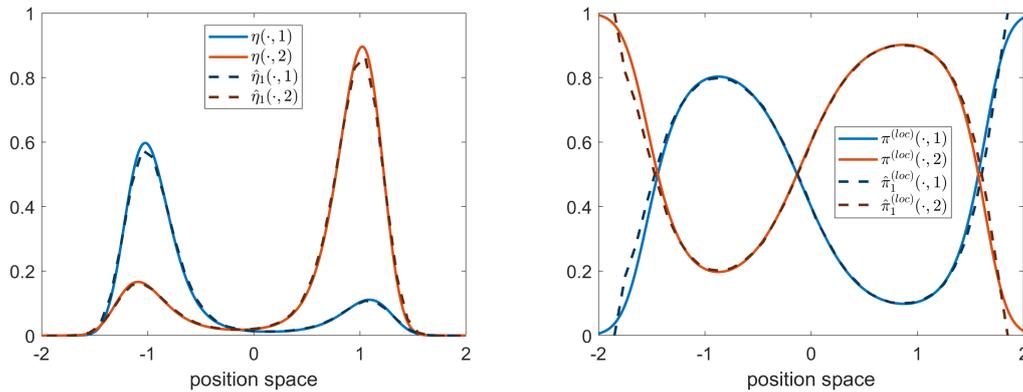


Figure 1.12: Example 7 with  $\pi$ -global information dynamics defined by the adoption rate functions in Equation (1.13): Empirical system state distributions  $\hat{\eta}_1(\cdot, 1), \hat{\eta}_1(\cdot, 2)$  (left) and empirical local status distributions  $\hat{\pi}_1^{(\text{loc})}(\cdot, 1), \hat{\pi}_1^{(\text{loc})}(\cdot, 2)$  (right) for agents are plotted with dashed lines. For comparison the respective theoretical distributions  $\eta(\cdot, 1), \pi^{(\text{loc})}(\cdot, 1)$  and  $\eta(\cdot, 2), \pi^{(\text{loc})}(\cdot, 2)$  are plotted as well. The curves referring to status 1 are marked blue and the curves referring to status 2 are marked orange.

empirical equilibrium state  $\hat{\eta}_1$  from the numerical example (see Figure 1.12) is very close ( $d_{\text{TV}}(\eta, \hat{\eta}_1) \approx 0.005$ ) to the theoretical equilibrium given by the Boltzmann distributions for each diffusion process weighted with  $\pi$ . This result suggests that the number of Monte Carlo simulations used for sampling is high enough for our purposes.

If instead we want to achieve the target distribution  $\eta$ , but only with local information within a radius  $r$  available to each agent, it is reasonable to define the adoption dynamics so that with increasing radius  $r \rightarrow \infty$  we return to the global information case defined by  $g_{ij}^{(\alpha)}$  in Equation (1.13). One approach to obtaining this property would be to define the rate functions as before, but using the population statistics of a sample from the local neighborhood of agent  $\alpha$  instead of the global population statistics.

### $\pi$ -Local Information

Let agent  $\alpha$  be in status  $i$ ,  $\gamma > 0$  a constant for scaling the jump intensity and  $r > 0$  the interaction radius, which determines the sampling area around the position  $x_\alpha$  of an

agent. Then, we define the adoption rate function as

$$f_{ij}^{(\alpha)}(X, S, t) := \gamma \max \left\{ 0, \left( \pi_j - \frac{\sum_{\beta=1}^{n_a} d_r(x_\alpha, x_\beta) \delta_j(s_\beta)}{\sum_{\beta=1}^{n_a} d_r(x_\alpha, x_\beta)} \right) \right\} \quad (1.15)$$

where  $d_r$  is the distance indicator function defined in (1.9) and  $\delta_j$  is the indicator function of status  $j$ . The adoption rate depends on the local population statistics and increases linearly from 0 if the fraction of agents in status  $j$  in the local sample is less than the value  $\pi_j$  of the given status distribution.

With this definition of adoption dynamics, agents change their status to reach  $\pi$  in the status distribution not globally but locally. This leads in general to a different equilibrium for the system state than with the dynamics of  $\pi$ -global information defined in (1.13), since the local status distribution and thus also the spatial distribution of agents in a given status is different. Although it is clear that  $f_{ij}^{(\alpha)} \rightarrow g_{ij}^{(\alpha)}$  for  $r \rightarrow \infty$ , the equilibrium of this system is not necessarily  $\eta$  for a sufficiently small radius  $r$ .

If we look at the agent system of our numerical Example 7 we have the case of a different equilibrium distribution. If we consider the master equation for a single agent that is part of an infinite population distributed according to  $\eta$ , we would get  $\pi^{(\text{loc})}(x, 1), \pi^{(\text{loc})}(x, 2)$  as values for the local population sample centered around position  $x$ , which would lead to the following equation:

$$\begin{aligned} \partial_t \eta(x, 1) &= \pi_1 L_1 \mu_1 + \pi_2 L_2 \mu_2 \\ &\quad - \pi_1 \int_{\mathbb{X}} f_{12}^{(\alpha)}(x) d\mu_1(x) + \pi_2 \int_{\mathbb{X}} f_{21}^{(\alpha)}(x) d\mu_2(x) \\ &= 0 + 0 - \pi_1 \int_{\mathbb{X}} \gamma \max\{0, \pi_2 - \pi^{(\text{loc})}(x, 2)\} d\mu_1(x) \\ &\quad + \pi_2 \int_{\mathbb{X}} \gamma \max\{0, \pi_1 - \pi^{(\text{loc})}(x, 1)\} d\mu_2(x). \end{aligned} \quad (1.16)$$

By the construction of our example this difference is nonzero, so the change of the probability density  $\eta$  is nonzero, and therefore  $\eta$  can not be the equilibrium density of the dynamics defined in Equation (1.15).

As expected, the dynamics with  $\pi$ -local information leads to an empirical equilibrium  $\hat{\eta}_2$  that is not close to  $\eta$  ( $d_{\text{TV}}(\eta, \hat{\eta}_2) \approx 0.1364$ ). The status distribution also deviates globally ( $\hat{\pi}_2 \approx (0.42, 0.58)$ ) as well as locally from  $\pi$ , which is not quite expected since the adoption rate functions are constructed in such a way that the status distribution should converge to  $\pi$  at least locally. The deviation is due to the feedback between adoptions and mobility, as agents are more attracted to different wells depending on their status. Figure 1.13 visualizes the distributions of this scenario.

So if we want the system to converge to the same target distribution  $\eta$  as in the case of

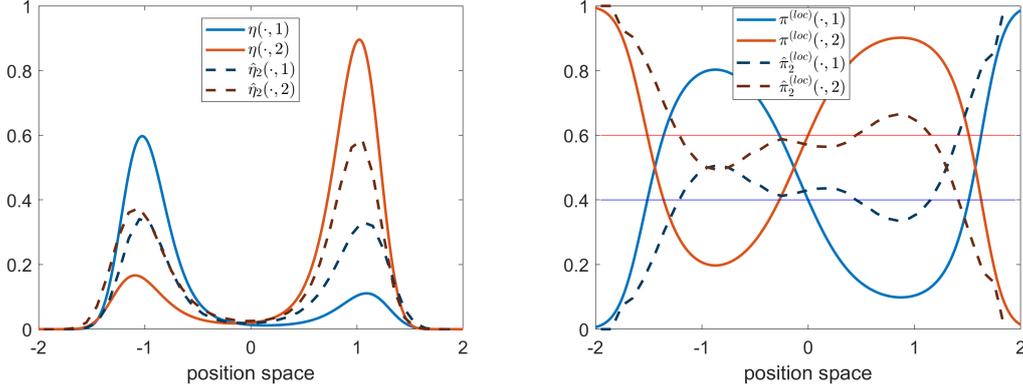


Figure 1.13: Example 7 with  $\pi$ -global information dynamics defined by the adoption rate functions in Equation (1.15): Empirical system state distributions  $\hat{\eta}_2(\cdot, 1), \hat{\eta}_2(\cdot, 2)$  (left) and empirical local status distributions  $\hat{\pi}_2^{(\text{loc})}(\cdot, 1), \hat{\pi}_2^{(\text{loc})}(\cdot, 2)$  (right) for agents are plotted with dashed lines. For comparison the respective theoretical distributions  $\eta(\cdot, 1), \pi^{(\text{loc})}(\cdot, 1)$  and  $\eta(\cdot, 2), \pi^{(\text{loc})}(\cdot, 2)$  are plotted as well. The curves referring to status 1 are marked blue and the curves referring to status 2 are marked orange. The blue and red horizontal lines mark the status distribution  $\pi$  which the  $\pi$ -local information dynamics try to achieve locally.

$\pi$  global information, but with only the information of a local neighborhood available to the agents, we need a different approach.

Our second approach to solving this problem requires the utilization of the function for the local status distribution  $\pi^{(\text{loc})}$ . From the system of Equations (1.16) it becomes clear how to modify the adoption rate functions such that  $\eta$  emerges as the stationary distribution. Instead of comparing the local status distribution from the sample of an agent  $\alpha$  to the global status distribution  $\pi$  for determining the adoption rate, we should compare it to the averaged local status distribution

$$\frac{\int_{B_r(x_\alpha)} \pi^{(\text{loc})}(x, \cdot) dx}{\lambda(B_r(x_\alpha))} \quad (1.17)$$

within a ball  $B_r$  with radius  $r$ . For  $r \rightarrow \infty$  this would be the same as the comparison with  $\pi$ , so with increasing radius we would converge to the case of agents with global information. On the other hand, we can assume that for a sufficiently small radius  $r$  it holds that

$$\frac{\int_{B_r(x_\alpha)} \pi^{(\text{loc})}(x, \cdot) dx}{\lambda(B_r(x_\alpha))} \approx \pi^{(\text{loc})}(x_\alpha, \cdot).$$

So, we can replace (1.17) with  $\pi^{(\text{loc})}$  in the definition of the adoption rate function to avoid calculating numerical integrals in the simulation steps.

### $\pi^{(\text{loc})}$ -Local Information

Let agent  $\alpha$  be in status  $i$ ,  $\gamma > 0$  a constant for scaling the jump intensity and  $r > 0$  the interaction radius, which determines the sampling area around the position  $x_\alpha$ . Then, we define the adoption rate function as

$$\ell_{ij}^\alpha(X, S, t) := \gamma \max \left\{ 0, \left( \pi^{(\text{loc})}(x_\alpha, j) - \frac{\sum_{\beta=1}^{n_a} d_r(x_\alpha, x_\beta) \delta_j(s_\beta)}{\sum_{\beta=1}^{n_a} d_r(x_\alpha, x_\beta)} \right) \right\} \quad (1.18)$$

with  $\pi^{(\text{loc})}$  as defined in (1.14),  $d_r$  the distance indicator for radius  $r$  and  $\delta_j$  the indicator function for status  $j$ .

The adoption rate depends on the local population statistics and now increases linearly from 0 if the fraction of agents with status  $j$  in the sample of agent  $\alpha$  is smaller than the value  $\pi^{(\text{loc})}(x_\alpha, j)$  of the given local status distribution.

By construction the target equilibrium density  $\eta$  emerges from the agent system equipped with these dynamics. Since the observed sample of agents in status  $j$  for every agent  $\alpha$  converges to  $\pi^{(\text{loc})}(x_\alpha, j)$  and therefore the rate functions  $\ell_{ij}^\alpha$  are zero for a population that is distributed according to  $\eta$ . Also, the change in the position part of  $\eta$  is zero, since it is constructed from the invariant densities of the generators of the diffusion processes for the agent mobility.

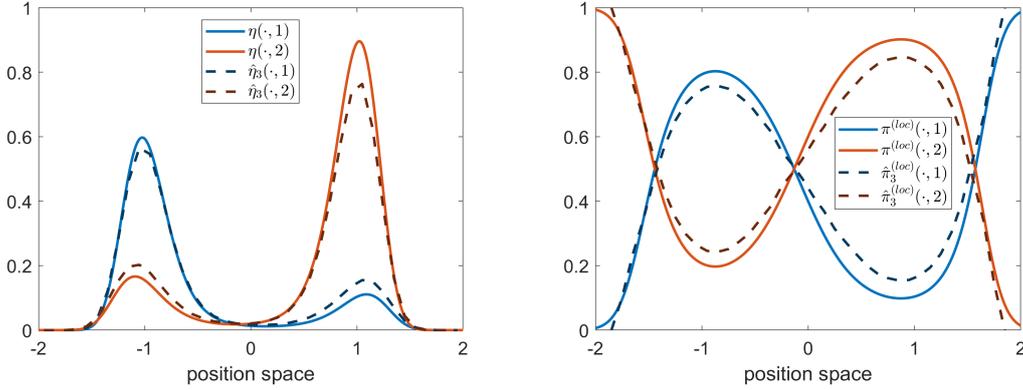


Figure 1.14: Example 7 with  $\pi^{(\text{loc})}$ -local information dynamics defined by the adoption rate functions in Equation (1.18). Empirical system state distributions  $\hat{\eta}_3(\cdot, 1), \hat{\eta}_3(\cdot, 2)$  (left) and empirical local status distributions  $\hat{\pi}_3^{(\text{loc})}(\cdot, 1), \hat{\pi}_3^{(\text{loc})}(\cdot, 2)$  (right) for agents are plotted with dashed lines. For comparison the respective theoretical distributions  $\eta(\cdot, 1), \pi^{(\text{loc})}(\cdot, 1)$  and  $\eta(\cdot, 2), \pi^{(\text{loc})}(\cdot, 2)$  are plotted as well. The curves referring to status 1 are marked blue and the curves referring to status 2 are marked orange.

The empirical equilibrium state of the numerical example under  $\pi^{(\text{loc})}$ -local information

dynamics  $\hat{\eta}_3$  is a good approximation for the equilibrium of the  $\pi$ -global information dynamics ( $d_{\text{TV}}(\eta, \hat{\eta}_3) \approx 0.0215$ ), although again there is a small deviation from the local status distribution that the system tries to achieve by construction. Unfortunately, also in this scenario there is a deviation from the global status distribution as we have  $\hat{\pi}_3 \approx (0.42, 0.58)$ . The convergence to  $\eta$  can be improved by increasing the number of agents or the interaction radius compared to the chosen values in this realization. This is due to a better estimation of the observed local status distribution through a larger local sample. The speed of the adoption dynamics is also crucial for the convergence to the target equilibrium  $\eta$  and we have already chosen it fast enough such that we are in the regime where further increasing the jump intensity will not significantly improve the results.

A possible application area for this modeling approach could be the field of systems biology, as many phenomena of biological organization depend on local interactions [50]. Especially for the process of morphogenesis, it is assumed that cells adapt according to local spatial information [51]. The adoption dynamics defined in Equation (1.18) is one way to encode positional information in a reaction-diffusion type ABM so that agents adapt their status according to an observed concentration. As the relationship between reaction-diffusion models and models based on positional information for morphogenesis is still being discussed [52], an adaptation of our example to a real-world system could be an interesting topic for future research.

### 1.3 Simulation of Agent-Based Models

As the dynamics of an agent-based model is usually defined through a set of stochastic differential equations which are not analytically solvable, we usually need to compute a numerical solution for the evaluation of the ABM. Since the ABMs we consider in general include randomness, we need to realize multiple Monte Carlo simulations for generating a sufficiently large sample to gather statistics about the distributions of our quantities of interest, e.g., first hitting times for critical transition events. Besides costly function evaluations, e.g., in the case of neighborhood computations, the generation of pseudorandom numbers is another factor which significantly contributes to the overall computational effort. Therefore, there will always be the trade-off between additional accuracy for single numerical realizations of the model and the realization of additional simulation runs to increase the sample size. Using efficient algorithms for the simulation of ABMs can thus allow us to make the best use out of limited computational resources.

In this section we will briefly review some well-known established numerical discretization schemes for stochastic processes. We will first discuss the numerical realization of the Brownian motion and Markov diffusion processes and afterwards the simulation of jump processes. We will then proceed with combined simulation approaches for jump diffusion processes and how to apply the algorithms to the ABMs that we are interested in. An event-based approach [15] that is well suited to simulate processes that possess a low jump intensity will be illustrated at the end of the section with a numerical example.

### 1.3.1 Simulation of Diffusion Processes

As is also the case for deterministic integrals and for that matter any continuous object, we will never be able to construct a discrete object that fully captures all properties of its continuous counterpart. We can only hope to get an approximation that is accurate enough in the sense that the error that we make when evaluating the discrete approximation instead is small and that the key properties are maintained.

When it comes to Brownian motion we know that the process has independent Gaussian increments. So, for all times  $t$  and all time-step sizes  $\Delta t \in \mathbb{R}$  it holds that

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

The sequence of random variables defined by

$$Z_{n+1} = Z_n + \sqrt{\Delta t} \xi_n \tag{1.19}$$

with  $Z_0 = 0$ ,  $\Delta t > 0$  the time step size and  $\xi_n \sim \mathcal{N}(0, 1)$  independent for all  $n \in \mathbb{N}$  has by construction also independent Gaussian increments and each  $Z_n$  is distributed as a Brownian motion at the same point in time would be ( $Z_n \sim \mathbf{B}(n\Delta t)$ ) in the case of a constant step size  $\Delta t$ ). However, between the discrete time points the process defined by the  $Z_n$  and linear interpolation inbetween behaves vastly different than a Brownian motion. Only for a sufficiently small time step size  $\Delta t$  the approximation error will be small [53]. If we view a sequence  $0 < t_1 < \dots < t_n$  as a discretization of the Lebesgue measure on  $\mathbb{T} \subseteq \mathbb{R}$  that measures each interval with  $\Delta t$ , then we can view the sequence of scaled Gaussian random variables defined by  $\sqrt{\Delta t} \xi_n$  as a discretization of the Gaussian random measure generated by the Brownian motion for the same intervals.

#### Euler-Maruyama scheme

To create a realization of a diffusion process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  defined by a stochastic differential equation of the type

$$d\mathbf{X}(t) = b(\mathbf{X}(t), t)dt + \sigma(\mathbf{X}(t), t)d\mathbf{B}(t)$$

one has to both solve the deterministic integral and realize the stochastic integral with respect to Brownian motion numerically. A well-known method to compute this is the Euler-Maruyama scheme which combines the explicit forward Euler method with an added properly scaled discrete approximation of the Brownian motion [53]. The Euler-Maruyama scheme can be defined as

$$\hat{\mathbf{X}}(t + \Delta t) = \hat{\mathbf{X}}(t) + b(\hat{\mathbf{X}}(t), t)\Delta t + \sigma(\hat{\mathbf{X}}(t), t)\sqrt{\Delta t}\xi_t \tag{1.20}$$

with step size  $\Delta t$  and  $\xi_t \sim \mathcal{N}(0, 1)$  independent random variables. The step size  $\Delta t$  can be a random variable as well, as long as there is a strict upper bound given by a positive constant  $\delta > \Delta t$  for all  $t$  [53]. The Euler-Maruyama scheme has weak convergence order 1 and strong convergence order 0.5 [53].

## Higher order methods and splitting schemes

Though there also exist higher order schemes based on explicit strong Itô-Taylor approximations [53], e.g., the Milstein scheme, the Euler-Maruyama scheme is still widely used. Even though the order of convergence is weaker, the effort compared to higher order methods is considerably less as it requires less function evaluations and less draws of pseudorandom numbers. For the sampling of most quantities of interest weak convergence of the sample paths already is sufficient and the bottleneck lies then rather in the number of realizations  $N$  as the Monte Carlo sampling converges only with  $\sqrt{N}$  to the expectation.

Recently also splitting methods have been proposed to improve the convergence order without including higher order terms [54]. Leimkuhler and Matthews have proposed an adaption to the Euler-Maruyama scheme that can be written down as

$$\hat{\mathbf{X}}(t_{n+1}) = \hat{\mathbf{X}}(t_n) + b(\hat{\mathbf{X}}(t_n), t_n)\Delta t + \sigma \frac{\sqrt{\Delta t}}{2}(\xi_{t_n} + \xi_{t_{n+1}})$$

with  $t_{n+1} := t_n + \Delta t$  a sequence of time points and  $\xi_{t_n} \sim \mathcal{N}(0, 1)$  independent standard normal random variables. The drawback of the method is that the increments are no longer independent but correlated with each other due to the reuse of one previous random variable in each time step. The process can however be made Markovian again by expanding the state space with the random variable that is reused in the next iteration step [54].

### 1.3.2 Simulation of Jump Processes

A time-continuous Markov chain on a finite state space  $\mathbb{S}$  with rate matrix  $\Lambda$  can be discretized by calculating a transition matrix  $P^{\Delta t}$  for a chosen time step size  $\Delta t$  by evaluating  $\int_0^{\Delta t} \exp(t\Lambda) dt$ . However, for large sets  $\mathbb{S}$  the matrix multiplication is not well-conditioned, and for Markov chains that are metastable we will have holding probabilities close to 1 for appropriately small step sizes  $\Delta t$  which can result in a high simulation effort and numerical instability. Thus, also in this case a simulation approach that is based on the view of Markov chains in terms of Poisson processes is useful.

As in the case of diffusion processes the first step is to construct a discretization of the Brownian motion, it is natural that for the simulation of jump processes we are interested in a discretization of the (unit rate) Poisson process. In the Poisson case we have the advantage compared to Brownian motion that the state changes due to random events already only happen at discrete time points for the original object and not continuously. For the realization of a homogeneous counting process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  with jump intensity  $\lambda > 0$  there are two common approaches.

#### Time discretization approach

The first approach would be analogous to the Brownian motion case: We fix a time step  $\Delta t$  for the next iteration of the process and use the property of independent increments. Since we have

$$\mathbf{N}(t + \Delta t) - \mathbf{N}(t) \sim \text{Pois}(\lambda \Delta t)$$

for the original process the iterative definition that preserves this quality is the process defined by

$$\hat{\mathbf{N}}(t + \Delta t) = \hat{\mathbf{N}}(t) + \rho_t \tag{1.21}$$

with  $\hat{\mathbf{N}}(0) = 0$ ,  $\rho_t \sim \text{Pois}(\lambda \Delta t)$  independent and constant interpolation between the evaluations, which by construction also has independent increments that are Poisson distributed with the same intensity  $\lambda$  as the original process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$ . With this construction we again have a good approximation for sufficiently small time steps  $\Delta t$  in the sense that the evaluations at each time of the sequence  $0 < t_1 < \dots < t_n$  are correctly distributed but inbetween we again have a deviation from the original characteristics of the process. We can again interpret the sequence of Poisson random variables that is induced by the time discretization as a discretization of the Poisson random measure that is generated by the counting process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$ .

#### Event-based approach

The second approach would be to simulate each state change event of  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  separately. As we have an exponentially distributed waiting time  $\tau \sim \text{Exp}(\lambda)$  between two state change events it suffices to draw a sequence of exponential random variables  $\tau_1, \dots, \tau_n \sim \text{Exp}(\lambda)$  and define iteratively

$$\hat{\mathbf{N}}(t + \tau_k) = \hat{\mathbf{N}}(t) + 1 \tag{1.22}$$

with  $\hat{\mathbf{N}}(0) = 0$ . This approach has the advantage that it provides an essentially exact realization of  $(\mathbf{N}(t))_{t \in \mathbb{T}}$ . However, in the case of a high intensity  $\lambda$  we will have to draw significantly more pseudorandom numbers compared to the first approach.

Both approaches can easily be adapted to the case of homogeneous compound Poisson processes  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  with Markov kernel  $\mathcal{Q}$ . In the case of the event-based approach we draw for each waiting time  $\tau_k$  also a random variable  $Z_k \sim \mathcal{Q}$  and adapt the definition from (1.22) to

$$\hat{\mathbf{X}}(t + \tau_k) = \hat{\mathbf{X}}(t) + Z_k. \quad (1.23)$$

In the case of the time discretization approach we draw a total of  $\rho_t$  random variables  $Z_k \sim \mathcal{Q}$  for each time step and adapt the increments in (1.21) accordingly to

$$\hat{\mathbf{X}}(t + \Delta t) = \hat{\mathbf{X}}(t) + \sum_{k=1}^{\rho_t} Z_k. \quad (1.24)$$

### Stochastic simulation for inhomogeneous jump processes

In the case of an intensity rate function  $\lambda$  we need to calculate a numerical integral to determine the jump intensity correctly for either drawing correctly distributed waiting times  $\tau \sim \text{Exp}(\int \lambda dt)$  or a correctly distributed number of jump events. Here we can distinguish between the two cases of a dependency of the intensity rate function purely on the state of the jump process, which does not change between two jump events, and an either direct or indirect dependence on time itself.

#### State-dependent jump intensity

In the first case scenario the value  $\lambda(\mathbf{X}(t))$  is constant between two jump events and thus for calculating the jump intensity between two iteration steps we only need to evaluate the rate function  $\lambda(\hat{\mathbf{X}}(t))$  at the value of the last iteration. We then can draw in the event-based discretization approach an exponential waiting time  $\tau \sim \text{Exp}(\lambda(\hat{\mathbf{X}}(t)))$  to determine the time  $t + \tau$  of the next iteration and draw a random variable  $Z$  according to the Markov kernel  $\mathcal{Q}(\hat{\mathbf{N}}(t))$ . This method is again essentially exact and also known as the stochastic simulation algorithm [55].

For the time discretization approach we need to assume that the time step size  $\Delta t$  is small enough such that the intensity rate function  $\lambda$  is almost constant between  $t$  and  $t + \Delta t$ , i.e., the state changes between two evaluations of the discretization scheme do not have a strong influence on the jump intensity [56]. We then can draw the number of jump events  $\rho_t \sim \text{Pois}(\lambda(\hat{\mathbf{X}}(t))\Delta t)$  for the system state update after time  $\Delta t$  and the random variables  $Z_k$  with respect to the Markov kernel  $\mathcal{Q}(\hat{\mathbf{X}}(t))$  for the update of the system. Depending on the definition of  $\mathcal{Q}$ , we might need to update the discretization of the Markov kernel after the draw of each  $Z_k$ , e.g., by considering  $Z_{k+1} \sim \mathcal{Q}(\hat{\mathbf{X}}(t) + Z_k)$ , to avoid a sequence of events that is dynamically impossible for the original process (e.g.,

mutually exclusive events that both have a positive intensity rate, such as an agent doing an  $i \rightarrow j_1$  transition as well as an  $i \rightarrow j_2$  transition at the same time for  $j_1 \neq j_2$ ). Alternatively, an update through such a combination of  $Z_k$  could be avoided by a form of rejection sampling. The method described by this time discretization approach is also known as tau-leaping in the literature [56]. Again, this approach is mostly useful in the case of a generally high jump intensity as we then gain more efficiency compared to the exact event-based algorithm. There are also combinations of both methods, where depending on the current jump intensity either the tau-leaping method or the stochastic simulation algorithm is used [57].

### Time-dependent jump intensity

In this case we have  $\lambda(\mathbf{X}(t), t)$  also changing between two jump events through the additional dependence on the time  $t$ . In the time discretization approach basically nothing changes compared to the previous case as we already needed to assume that  $\Delta t$  is chosen such that the assumption that the intensity rate function  $\lambda$  is almost constant on  $[t, t + \Delta t)$  holds.

For the event-based simulation approach we now also need to compute the numerical integral  $\int_t^{t+\tau} \lambda(\mathbf{X}(s), s) ds$  such that we can draw a correctly distributed waiting time  $\tau$  between two events. We know that we can construct an inhomogeneous Poisson process with intensity rate function  $\lambda$  by scaling the time of a unit rate Poisson process according to  $\lambda$ . Thus, we can draw a waiting time  $\tau' \sim \text{Exp}(1)$ , which is the waiting time for a unit rate Poisson process, in advance and by solving

$$\tau' = \operatorname{argmin}_{\tau} \left\{ \int_t^{t+\tau} \lambda(\mathbf{X}(s), s) ds \right\}$$

we can calculate a correctly distributed waiting time  $\tau$ . An algorithmic way to realize this is presented in [58] for the case of a jump process on a time dependent network for modeling disease spreading. Translated to the general case of an inhomogeneous compound Poisson process the algorithm can be written as is done in Algorithm 1. Besides the numerical error from integrating the intensity rate function the jump times are stochastically exact.

This event-based simulation method is the algorithm of choice whenever the assumption that  $\lambda$  is almost constant on  $[t, t + \Delta t)$  holds only for sufficiently small  $\Delta t$  that in consequence lead then on average to jump events only happening after multiple time steps, i.e., the Poisson random variables drawn in the time discretization approach being 0 with a high probability.

The algorithm realizes the numerical integration of the intensity rate function  $\lambda$  through an explicit 1-step forward Euler scheme. While in principle the method could be adapted to utilize a higher order scheme for calculating the jump intensity, as is e.g., done in higher order tau-leaping methods [59], it then becomes unclear how to calculate the

adaptive step size in the jump event steps. A possible solution could be to use the forward Euler scheme to determine whether a jump event takes place and only apply a higher order scheme to improve the accuracy for time steps without jumps. Exploring whether accuracy can be gained by such an adaption of the event-based approach could be a topic for future research.

---

**Algorithm 1:** Event-based simulation of Compound Poisson processes

---

```

1 initialize time  $t = 0$  and state  $\hat{\mathbf{X}}(0) = \mathbf{X}(0)$ ;
2 choose a time step  $\Delta t$  and time horizon  $T$ ;
3 draw  $\tau' \sim \text{Exp}(1)$  # exponentially distributed with rate 1;
4 while  $t < T$  do
5     # compute the jump intensity rate;
6     set  $\Lambda := \lambda(\hat{\mathbf{X}}(t), t)$ ;
7     if  $\Lambda \Delta t > \tau'$  then
8         # jump event;
9         draw  $Z$  according to the Markov kernel  $\mathcal{Q}(\hat{\mathbf{X}}(t), t)$ ;
10        # state update;
11         $\hat{\mathbf{X}}\left(t + \frac{\tau'}{\Lambda}\right) = \hat{\mathbf{X}}(t) + Z$ ;
12        # time update;
13         $t = t + \frac{\tau'}{\Lambda}$ ;
14        draw new  $\tau' \sim \text{Exp}(1)$ ;
15    else
16        # time update;
17         $t = t + \Delta t$ ;
18         $\tau' = \tau' - \Lambda \Delta t$ ;
19    end
20 end
Result:  $(\hat{\mathbf{X}}(t))_{t \leq T}$ 

```

---

### 1.3.3 Combined Simulation Approach

We will now combine the simulation methods for realizing jump processes and diffusion processes to get a simulation scheme for jump diffusion processes and subsequently agent-based models that can be formalized as Markov processes based on jump diffusions.

We will expand the event-based Algorithm 1 step by step and adapt it first for the simulation of PDMPs and afterwards for the simulation of jump diffusions. In the case of a consistently high jump intensity, a similar adaptation to the time discretization method for jump processes might perform better, but as the adaptation is pretty straightforward, we will in the following only discuss the event-based approach in detail.

#### Simulation of PDMPs

In the case of  $(\mathbf{Y}(t))_{t \in \mathbb{T}}$  being a piecewise-deterministic Markov process given by

$$d\mathbf{Y}(t) = b(\mathbf{Y}(t), t)dt + Z_{\mathbf{N}(t)}d\mathbf{N}(t)$$

with Markov kernel  $\mathcal{Q}$  and  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  a Poisson process with intensity rate function  $\lambda$ , we have the situation that the state of the process also changes between two jump events so we have at least an implicit time dependence of the intensity rate function through the deterministic development of the process. Thus, we adapt Algorithm 1 for event-based simulation of inhomogeneous compound Poisson processes by combining it with an explicit forward Euler scheme to update the state of the process between jump events according to the deterministic drift term  $b$ .

Again, in principle we could also utilize a higher order method for all fixed step size iteration steps without jump events and use the forward Euler method only for the iterations of the process in which a jump event is calculated and the time step size adapted.

The corresponding time discretization approach to simulating a PDMP can be viewed as the analogue to the Euler-Maruyama scheme for SDEs driven by a deterministic drift and Poisson noise. As there are higher order tau-leaping methods one could also utilize such a scheme for the simulation of PDMPs.

---

**Algorithm 2:** Event-based simulation of PDMPs

---

```
1 initialize time  $t = 0$  and state  $\hat{\mathbf{X}}(0) = \mathbf{X}(0)$ ;  
2 choose a time step  $\Delta t$  and time horizon  $T$ ;  
3 draw  $\tau' \sim \text{Exp}(1)$  # exponentially distributed with rate 1;  
4 while  $t < T$  do  
5   # compute the jump intensity rate;  
6   set  $\Lambda := \lambda(\hat{\mathbf{X}}(t), t)$  ;  
7   if  $\Lambda \Delta t > \tau'$  then  
8     # jump event;  
9     draw  $Z$  according to the Markov kernel  $\mathcal{Q}(\hat{\mathbf{X}}(t), t)$ ;  
10    # state update;  
11     $\hat{\mathbf{X}}\left(t + \frac{\tau'}{\Lambda}\right) = \hat{\mathbf{X}}(t)b\left(\hat{\mathbf{X}}(t), t\right) \frac{\tau'}{\Lambda} + Z$ ;  
12    # time update;  
13     $t = t + \frac{\tau'}{\Lambda}$ ;  
14    draw new  $\tau' \sim \text{Exp}(1)$ ;  
15  else  
16    # state update;  
17     $\hat{\mathbf{X}}(t + \Delta t) = \hat{\mathbf{X}}(t) + b\left(\hat{\mathbf{X}}(t), t\right) \Delta t$ ;  
18    # time update;  
19     $t = t + \Delta t$ ;  
20     $\tau' = \tau' - \Lambda \Delta t$ ;  
21  end  
22 end  
  Result:  $(\hat{\mathbf{X}}(t))_{t \leq T}$ 
```

---

### Simulation of jump diffusion processes

Finally, for the simulation of jump diffusion processes  $(\mathbf{Y}(t))_{t \in \mathbb{T}}$  defined by an SDE of the form

$$d\mathbf{X}(t) = b(\mathbf{X}(t), t)dt + \sigma(\mathbf{X}(t), t)d\mathbf{B}(t) + Z_{\mathbf{N}(t)}d\mathbf{N}(t)$$

we can realize the simulation by adapting Algorithm 2 used for the simulation of PDMPs by updating the process between the jump events with an Euler-Maruyama scheme instead of the forward Euler.

---

#### Algorithm 3: Event-based simulation of jump diffusion processes

---

```

1 initialize time  $t = 0$  and state  $\hat{\mathbf{X}}(0) = \mathbf{X}(0)$ ;
2 choose a time step  $\Delta t$  and time horizon  $T$ ;
3 draw  $\tau' \sim \text{Exp}(1)$  # exponentially distributed with rate 1;
4 while  $t < T$  do
5     # compute the jump intensity rate;
6     set  $\Lambda := \lambda(\hat{\mathbf{X}}(t), t)$ ;
7     if  $\Lambda \Delta t > \tau'$  then
8         # jump event;
9         draw  $Z$  according to the Markov kernel  $\mathcal{Q}(\hat{\mathbf{X}}(t), t)$ ;
10        # state update;
11        draw  $\xi \sim \mathcal{N}(0, 1)$ ;
12         $\hat{\mathbf{X}}\left(t + \frac{\tau'}{\Lambda}\right) = \hat{\mathbf{X}}(t)b\left(\hat{\mathbf{X}}(t), t\right) \frac{\tau'}{\Lambda} + \sigma\left(\hat{\mathbf{X}}(t), t\right) \sqrt{\frac{\tau'}{\Lambda}}\xi + Z$ ;
13        # time update;
14         $t = t + \frac{\tau'}{\Lambda}$ ;
15        draw new  $\tau' \sim \text{Exp}(1)$ ;
16    else
17        # state update;
18        draw  $\xi \sim \mathcal{N}(0, 1)$ ;
19         $\hat{\mathbf{X}}(t + \Delta t) = \hat{\mathbf{X}}(t) + b\left(\hat{\mathbf{X}}(t), t\right) \Delta t + \sigma\left(\hat{\mathbf{X}}(t), t\right) \sqrt{\Delta t}\xi$ ;
20        # time update;
21         $t = t + \Delta t$ ;
22         $\tau' = \tau' - \Lambda \Delta t$ ;
23    end
24 end

```

**Result:**  $(\hat{\mathbf{X}}(t))_{t \leq T}$

---

The convergence of the combined method follows from the convergence of the Euler-Maruyama scheme, which also allows for random but bounded time steps as the ones we have, and the convergence of the stochastic simulation algorithm. The weak and strong order of convergence is limited by the convergence order of Euler-Maruyama. The algorithm can easily be adapted to utilizing the earlier mentioned adaptation of

Euler-Maruyama proposed by Leimkuhler and Matthews [54] by drawing an additional random variable when initializing the system and adapting the state update step accordingly.

For the time discretization approach we adapt the Euler-Maruyama scheme with the additional draw of a Poisson distributed random variable  $\rho$  according to the jump intensity and the random variables  $Z_k$  to realize the jump event. This method is sometimes also called extended Euler-Maruyama scheme [60].

---

**Algorithm 4:** Time discretization approach for simulation of jump diffusion processes

---

```

1 initialize time  $t = 0$ ,  $Z_0 = 0$  and state  $\hat{\mathbf{X}}(0) = \mathbf{X}(0)$ ;
2 choose a time step  $\Delta t$  and time horizon  $T$ ;
3 while  $t < T$  do
4     # compute the jump intensity rate;
5     set  $\Lambda := \lambda(\hat{\mathbf{X}}(t), t)$ ;
6     # state update;
7     draw  $\xi \sim \mathcal{N}(0, 1)$ ;
8     draw  $\rho \sim \text{Pois}(\Lambda \Delta t)$ ;
9     if  $\rho > 0$  then
10        for  $k = 1 : \rho$  do
11            | draw  $Z_k$  according to the Markov kernel  $\mathcal{Q}(\hat{\mathbf{X}}(t) + \sum_{i=0}^{k-1} Z_i, t)$ ;
12            | end
13        end
14         $\hat{\mathbf{X}}(t + \Delta t) = \hat{\mathbf{X}}(t) + b(\hat{\mathbf{X}}(t), t) \Delta t + \sigma(\hat{\mathbf{X}}(t), t) \sqrt{\Delta t} \xi + \sum_{k=0}^{\rho} Z_k$ ;
15        # time update;
16         $t = t + \Delta t$ ;
17 end

```

**Result:**  $(\hat{\mathbf{X}}(t))_{t \leq T}$

---

Again, the event-based simulation approach is more accurate than the time discretization approach, but it requires in the case of a high jump intensity more computational effort.

In general, the consideration of higher order methods for simulating the diffusion process between jump events with a higher accuracy has the same drawbacks as in the case of pure diffusion processes. Nevertheless, for the time discretization approach also higher order methods for jump processes are used, especially in mathematical finance [61, 62]. Utilizing higher order methods in the event-based approach could be a topic for future research.

## Simulation of ABMs

Now we will apply the two simulation approaches for jump diffusion processes to our agent-based model formulation from Section 1.2.2. We will compare the accuracy of both approaches for a simple numerical example and see that indeed the event-based approach is more accurate in this case.

We consider a system of  $n_a$  agents that take values in the state space  $\mathbb{Y} := \mathbb{X} \times \mathcal{S}$  with  $X$  encoding all agents' positions and  $S$  encoding the status of all agents in the system state  $Y = (X, S)$ . For simplicity we assume that the process for the agent mobility can be written down as an SDE of the form

$$d\mathbf{X}(t) = -\nabla V(\mathbf{Y}(t), t)dt + \sigma(\mathbf{Y}(t), t)d\mathbf{B}(t)$$

but of course also more general processes can be realized with the algorithms that have been presented so far. For the adoption process we assume a finite number of  $Jn_a$  state change vectors  $v_j^{(\alpha)}$  with corresponding adoption rate functions  $f_j^{(\alpha)}$  which allows us an explicit construction of the Markov kernel  $\mathcal{Q}$  at the time of a jump event. The index  $j$  indicates the event type while the index  $\alpha$  denoted the agent ID.

The movements of the agents can then be simulated via the Euler-Maruyama scheme, and in every time step without jump events the positions of the agents are updated according to

$$X(t + \Delta t) = X(t) - \nabla V(\mathbf{Y}(t), t)\Delta t + \sigma(\mathbf{Y}(t), t)\sqrt{\Delta t}\xi, \quad (1.25)$$

where  $\xi$  is drawn from a standard normal distribution in  $\mathbb{R}^{dn_a}$  with  $d$  denoting the dimension of  $\mathbb{X}$ . For the Euler-Maruyama scheme to be consistent, the time step  $\Delta t$  needs to be chosen sufficiently small. As the jump intensity of the adoption process depends in general on the positions, also the adoption rate functions need to be evaluated again after the update of the position vector. We now provide the pseudo code for the event-based simulation algorithm that is illustrated in Figure 1.15 and afterwards the pseudo code of a tau-leaping method for ABMs.

## Event-Based Approach for ABMs

---

### Algorithm 5: Event-based simulation algorithm for ABMs

---

```

1 initialize time  $t = 0$  and system state  $\mathbf{Y}(0) = (\mathbf{X}(0), \mathbf{S}(0))$  for  $n_a$  agents;
2 choose a time step  $\Delta t$  and time horizon  $T$ ;
3 draw  $\tau' \sim \text{Exp}(1)$  # exponentially distributed with rate 1;
4 while  $t < T$  do
5     # calculate rates for adoption events
6     for  $j = 1 \dots J$  and  $\alpha = 1 \dots n_a$  compute the adaption rate function
7          $f_j^{(\alpha)}(\mathbf{Y}(t), t)$ ;
8     calculate total adoption rate  $\Lambda(t) = \sum_{j=1}^J \sum_{\alpha=1}^{n_a} f_j^{(\alpha)}(\mathbf{Y}(t), t)$ ;
9     if  $\Lambda(t) \Delta t > \alpha$  then
10         # choose adoption event
11         define Markov kernel  $\mathcal{Q}$  with  $\mathcal{Q}(v_j^{(\alpha)}) := \frac{f_j^{(\alpha)}(\mathbf{Y}(t), t)}{\Lambda}$ ;
12         draw a state change vector  $v_j^{(\alpha)}$  according to  $\mathcal{Q}$ ;
13         # status update
14          $\mathbf{S}(t + \frac{\tau'}{\Lambda}) = \mathbf{S}(t) + v_j^{(\alpha)}$ ;
15         # position update
16         draw  $\xi \sim \mathcal{N}(0, 1)$ ;
17          $\mathbf{X}(t + \frac{\tau'}{\Lambda}) = \mathbf{X}(t) - \nabla V(\mathbf{Y}(t), t) \frac{\tau'}{\Lambda} + \sigma(\mathbf{Y}(t), t) \sqrt{\frac{\tau'}{\Lambda}} \xi$ ;
18         # time update
19          $t = t + \frac{\tau'}{\Lambda}$ ;
20         draw new  $\tau' \sim \text{Exp}(1)$ ;
21     else
22         # position update
23         draw  $\xi \sim \mathcal{N}(0, 1)^{n_a}$ ;
24          $\mathbf{X}(t + \Delta t) = \mathbf{X}(t) - \nabla V(\mathbf{Y}(t), t) \Delta t + \sigma(\mathbf{Y}(t), t) \sqrt{\Delta t} \xi$ ;
25         # time update
26          $t = t + \Delta t$ ;
27          $\tau' = \tau' - \Lambda \Delta t$ ;
28     end
29 end

```

**Result:**  $(\mathbf{Y}(t))_{t \leq T}$

---

In the actual implementation there is no extra step to define a separate object for the Markov kernel determining the jump events as it is already implicitly defined through the calculation of the jump intensity rates which are stored in a vector. For the pseudo code however I think that it is more clear to write it down in this way. The same comment will apply to all future algorithm steps in which Markov kernels are explicitly calculated.

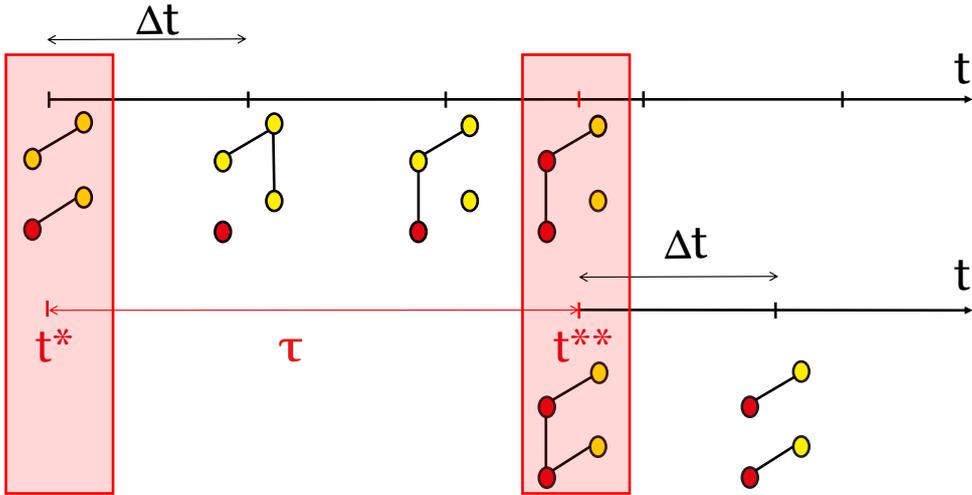


Figure 1.15: Illustration of Algorithm 5. The updates of the system state of an ABM are illustrated by a time-dependent interaction network that is implicitly defined through the second order adoptions in many applications. The colors of nodes represent the status and the edge indicates that two agents are closer than an interaction radius. Between two jump events at times  $t^*$  and  $t^{**}$  the positions of the agents are updated through Euler-Maruyama steps of size  $\Delta t$  and the time-dependent network changes accordingly.

For the realization of the jump event at time  $t^{**}$  we take a reduced step size such that the waiting time  $\tau$  that is the unique solution of

$$\tau' = \operatorname{argmin}_{\tau} \left\{ \int_{t^*}^{t^*+\tau} \sum_{j=1}^J \sum_{\alpha=1}^{n_a} f_j^{(\alpha)}(Y(t), t) dt \right\}$$

with  $\tau' \sim \operatorname{Exp}(1)$  being correctly realized. The jump event is illustrated by a status transition of the node from yellow to red. After the jump event at  $t^{**}$  we proceed with steps of size  $\Delta t$  until the next jump event is realized.

## Time Discretization Approach for ABMs

Pseudo code for a tau-leaping approach to simulating ABMs:

---

**Algorithm 6:** Time discretization approach for simulation of ABMs

---

```

1 initialize time  $t = 0$  and system state  $\mathbf{Y}(0) = (\mathbf{X}(0), \mathbf{S}(0))$  for  $n_a$  agents;
2 choose a time step  $\Delta t$  and time horizon  $T$ ;
3 while  $t < T$  do
4     # calculate rates for adoption events;
5     for  $k = 1 \dots J$  and  $\alpha = 1 \dots n_a$  compute the adoption rate function
         $f_k^{(\alpha)}(\mathbf{Y}(t), t)$ ;
6     calculate total adoption rate  $\Lambda(t) = \sum_{j=1}^J \sum_{\alpha=1}^{n_a} f_j^{(\alpha)}(\mathbf{Y}(t), t)$ ;
7     define Markov kernel  $\mathcal{Q}$  with  $\mathcal{Q}(v_j^{(\alpha)}) := \frac{f_j^{(\alpha)}(\mathbf{Y}(t), t)}{\Lambda}$ ;
8     # draw random variables;
9     draw  $\xi \sim \mathcal{N}(0, 1)^{n_a}$ ;
10    draw  $\rho \sim \text{Pois}(\Lambda \Delta t)$ ;
11    if  $\rho > 0$  then
12        for  $k = 1 : \rho$  do
13            | draw  $Z_k$  according to an adjusted Markov kernel  $\mathcal{Q}_k$ ;
14        end
15    end
16    # status update;
17     $\mathbf{S}(t + \Delta t) = \mathbf{S}(t) + \sum_{k=0}^{\rho} Z_k$ ;
18    #position update;
19     $\mathbf{X}(t + \Delta t) = \mathbf{X}(t) - \nabla V(\mathbf{Y}(t), t) \Delta t + \sigma(\mathbf{Y}(t), t) \sqrt{\Delta t} \xi$ ;
20    # time update;
21     $t = t + \Delta t$ ;
22 end
Result:  $(\mathbf{Y}(t))_{t \leq T}$ 

```

---

Depending on the adoption rate functions, in some cases it is reasonable to draw binomial random variables instead of Poisson random variables, e.g., to avoid negative populations in models for chemical reaction kinetics [57]. For a generally low jump intensity with a total adoption rate close to zero and/or a high population drawing Binomials is a good approximation to drawing Poisson random variables and can be more efficient [63]. If we assume that within a time step  $\Delta t$  an agent can only perform one action, then we can infer a probability  $p$  for the distribution  $B(n_a, p)$  such that  $\mathbb{E}(B(n_a, p)) = pn_a = \Lambda \Delta t = \mathbb{E}(\text{Pois}(\Lambda \Delta t))$ . This identity can also be used to translate ABMs that are defined in a more descriptive way with agents performing rejection sampling for their interactions in each time step to the corresponding formulation as a Poisson process. An action of a single agent that happens with probability  $p$  within a time step  $\Delta t$  is then related to a Poisson process with rate  $\lambda = \frac{p}{\Delta t}$ .

### Numerical Example: Single-Well Suitability Landscape

For illustration and comparison of the two simulation approaches, we consider a simpler version of our guiding Example 6 that was also presented in [15] by me and coauthors. We define a system with  $n = 20$  agents moving in a single-well potential landscape on  $\mathbb{X} = \mathbb{R}^2$  given by

$$V(x, y) = 10(x^2 + y^2)$$

with the status space being the set  $\mathbb{S} = \{0, 1\}$  with only second order adoptions for the transition from  $0 \rightarrow 1$ . The potential landscape has one local (and also global) minimum at  $(0, 0)$ , which for the agents corresponds to the most attractive point of the landscape. The diffusion process will then result in the agents' positions to be concentrated around the minimum. In order to study the spreading of status 1 among the agents, we initiate the status vector  $S$  of the agent system with  $s_1 = 1$  and  $s_\alpha = 0$  for all other agents  $\alpha \neq 1$ . With rate  $\gamma = 10$  the innovation spreads to other agents along the edges of the time-evolving contact network that arises from the evaluation of the distance indicator function  $d_r$  in the definition of the second order adoptions.

Next, we compare the two joint simulation approaches, i.e., the event-based simulation algorithm and the time discretization approach, which was realized for this example by a synchronous updating scheme [15], where in each time step and for each possible status transition a random number was drawn to determine whether the event is realized or not. Because of the very small time step and low number of agents, the total intensity rate of this system is so low that a rerun of this example with the tau-leaping method did not perform well, which is why we stick with the synchronous updating scheme in this case even though in general it is not recommended because of the additional effort. The results were averaged over a Monte Carlo sample of 2000 realizations of each parameter choice of  $\Delta t$ . In this example we choose  $\Delta t = 0.002$  as the largest time step size in order to obtain an accurate approximation of the agents' movements and the resulting numerical integral of the adoption rate functions. Given the time-evolving network based on the distance indicator function, for any  $\Delta t$  the event-based simulation approach provides statistically exact event times, whereas the synchronous updating scheme is a good approximation of the spreading process only for small  $\Delta t$  [64]. In order to compare the two combined approaches, we use the time discretization scheme with  $\Delta t = 0.0001$  as a benchmark. From the plot in Figure 1.16, one can see that the event-based simulation with  $\Delta t = 0.002$  indeed agrees with our benchmark. The synchronous updating for larger time steps  $\Delta t = 0.002$  deviates from the benchmark and on average the spreading is slower than the event-based approach as we have to draw multiple additional random numbers in each time step.

This example highlights again that a time discretization approach should only be considered for an agent system with a generally high jump intensity, as we have to choose a sufficiently small step size for realizing the diffusion part of the process but on the other hand tau-leaping does not perform well when we have an expected number of jump events within a time step is close to 0. So, unless the effective step size of the

event-based algorithm drops significantly lower than the default step size  $\Delta t$ , the event-based method is both more efficient and more accurate than the alternative. In this

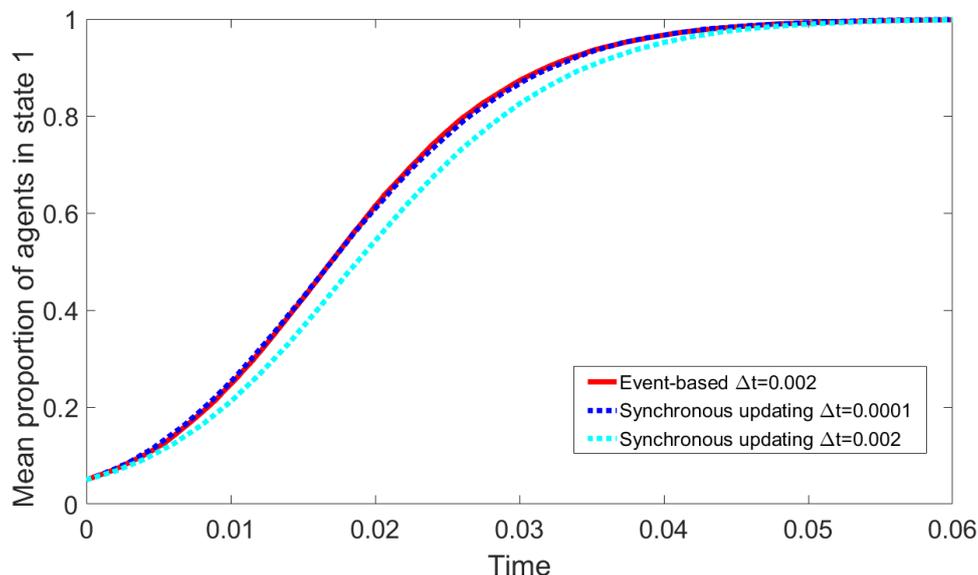


Figure 1.16: Illustration of the single-well example [15]. The dashed lines are showing the average adoption process within the well from a sample of 2000 simulations with the synchronous updating scheme for two different step sizes  $\Delta t$ . The result of the event-based simulation algorithm coincides with the benchmark of the synchronous updating scheme and has a greater accuracy than the synchronous updating scheme of the same step size.

example case, we have confirmed that the event-based approach is more accurate than the synchronous updating scheme for the same step-size, but also in other cases this holds [58, 64] due to the immediate response of the simulation algorithm to adaption events in the event-based approach. Besides the accuracy of the event-based scheme, we also have to consider computational costs. In general, the simulation time for the discretization of the diffusion process is the same for both simulation approaches as long as the same time step sizes are chosen. However, depending on the overall jump intensity for the system, each of the two approaches can be computationally advantageous. Especially in the case of jump events only happening at a fraction of the time steps, e.g., in the case of a spreading process with sparsely connected agents, the computational gain of the event-based algorithm is high. On the other hand, if jump events happen so frequently that the adaptive step size in the event-based approach leads to a highly increased number of time steps to simulate the same amount of time, then the synchronous updating scheme is significantly faster. In the application to models based on real-world data in Chapter 3 we will actually encounter both cases.

In the case of a large-scale agent system, the computational effort of both simulation approaches will be very high as the computational cost for evaluating the adoption rate functions usually scales non-linearly with the number of agents, e.g., because of neighborhood computations. So, a sufficiently large Monte Carlo sample of realizations will be at least very costly if not infeasible to calculate. Therefore, we discuss in the next chapter how to derive a reduced model that captures the essential properties of the ABM but takes significantly less effort to simulate.



## 2 Coarse Graining of Agent-Based Models

In the last chapter, we saw that Markov processes, although each of them has its own distinct characteristics, are fundamentally connected by their defining property of being memoryless. In many limiting cases one can be approximated through the other, e.g., by use of the central limit theorem or approximations through the expectation. In the modeling of chemical reaction kinetics a formulation as a jump process through the chemical master equation [65] can in limiting cases be approximated by a diffusion process given by the chemical Langevin equation [31] or even a purely deterministic process [66]. There are also hybrid approaches where the limiting procedure is applied to only a part of the dynamics and the system can then be modeled by a PDMP [30]. On the other hand, for metastable Markov processes that exhibit exponential exit time characteristics, we can derive an approximation by a discrete Markov chain through a suitable projection of the state space [13]. Each of the approximation steps leads to a reduced model, which loses some details of the original model in the approximation process, but in return requires significantly less simulation effort. The aim is of course to preserve the fundamental properties of the model that we are interested in for addressing the research question at hand. If we approximate e.g., an agent-based model for infection spreading by a compartmental ODE model then the spatial resolution is completely lost and some questions can no longer be answered by the coarser model.

In this chapter, we want to discuss a mesoscale approximation approach for agent-based models, so that we simplify the microscale dynamics but still keep some level of spatial resolution. We will utilize some of the aforementioned techniques for model reduction, namely the Markov state model (MSM) framework for projected dynamics on a reduced state space and the limiting approaches for jump processes with high intensity. We will in the process however lose the defining property of the agent-based model, as through the spatial coarse-graining the individual agents will no longer be distinguishable and in consequence the reduced model will be formulated as a metapopulation model [67]. We will consider two types of metapopulation models, one that has the dynamics of a pure jump process and as a second type, the piecewise-deterministic model formulation that is justified by population limit results [68]. The main argument for making this distinction is when a timescale separation is possible, such that some types of jump events have a significantly higher intensity rate than other types, so that already for an intermediate population we can apply the limit for the fast processes. The hierarchy of the different models and the connections through the approximation steps are illustrated in Figure 2.1.

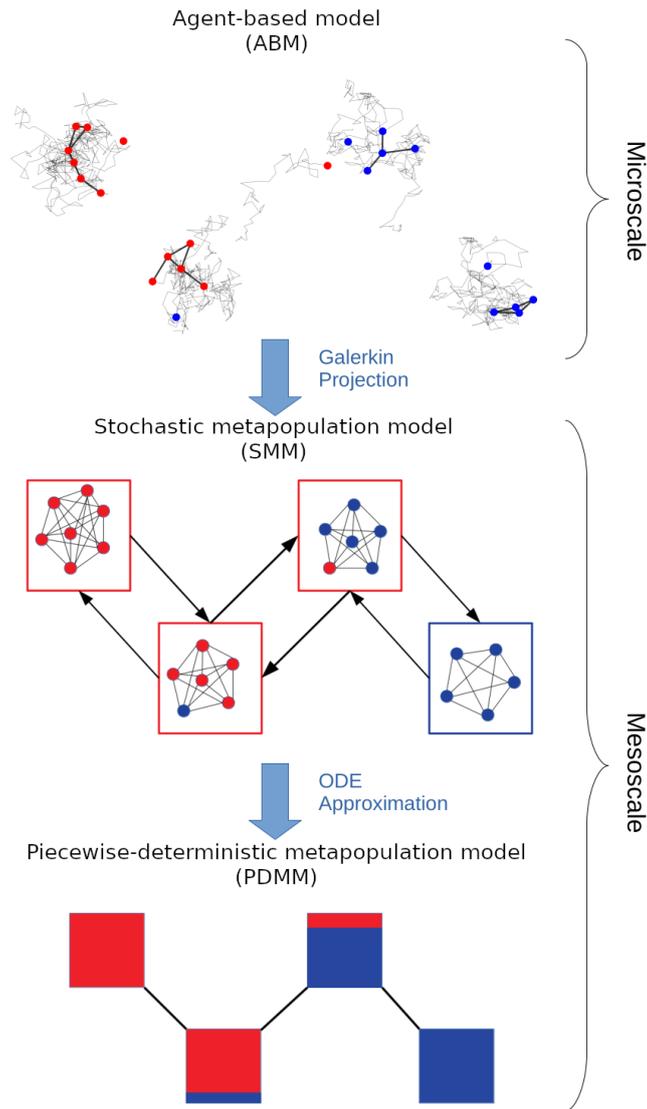


Figure 2.1: Illustration of the hierarchy of the modeling approaches considered in this chapter and the related steps of the coarse-graining procedure. Dots represent agents that have a status (blue or red), mesoscale boxes represent subpopulations. Straight lines between dots indicate connections between agents that allow interaction. Grey lines on the microscale show movement trajectories of the agents, arrows and lines between boxes on the mesoscale represent possible transitions between subpopulations. For the piecewise-deterministic metapopulation model, the population state is depicted by the proportion of blue and red.

It is of course natural to also consider a hybrid model that utilizes a diffusion approximation [31] of the fast jumps instead of a deterministic one and there might be applications where we also want to preserve some kind of stochasticity in the fast dynamics. The resulting model would then be a jump diffusion metapopulation model with the advantage that also for the fast dynamics some stochasticity is maintained but the disadvantage that it takes significantly more effort to simulate compared to the PDMP counterpart. However, in this work we will not discuss this in detail and leave it as a topic for future research.

We will now first recall the Markov state model framework for projected transfer operators presented in [13, 69], as some readers may not be familiar with it, and then discuss in detail our model reduction approach for ABMs that can be described as Markov processes. We will also illustrate the approach on the guiding example of the first chapter. The results on the stochastic and piecewise-deterministic metapopulation models were presented by me and co-authors in [16] and I will present them in a similar way in this work, including some literal passages from the publication.

## 2.1 Markov State Model Theory

A good motivation for the concept of Markov state models is the situation of the double-well potential example in Section 1.1.2. In Figure 1.4 we see that the trajectory concentrates around the two local minima of the double-well potential and takes values outside the two wells only during the rare transitions between the neighborhoods of the local minima. The exit time from one of the two wells can be approximated by an exponential waiting time, with the approximation quality depending on the diffusion coefficient and the steepness of the well. If we fix the diffusion coefficient, then the deeper the well is the longer the process will remain on average close to the minimum before exiting the neighborhood and the approximation of the exit time with an exponential random variable will be better. This property is called metastability and has been extensively studied analytically as well as computationally [13, 70]. Metastable systems can be characterized by multiscale dynamics, i.e., we can distinguish between dynamics on fast and slow time scales [13]. In the double-well example the diffusion within the neighborhood of an attractor would be considered a fast process compared to the much slower transition process between the wells. If we are only interested in which of the two wells the process is located at the time, but not in the exact position within one of the wells, then we can consider a simpler model that we can derive through a projection of the process according to a partition of the state space into two sets that each contain one of the two wells. As we have asymptotically exponential exit time characteristics the slow transition process between the two wells of the system can be approximated by a pure jump process that switches between two discrete states that represent the sets of the partition. The characteristics of the fast diffusion process within a well is, however, no longer present in the reduced model. The larger the time scale difference of these processes is, the lower the approximation error will be.

The Markov state model framework describes an approach how to derive such a reduced model  $(\hat{\mathbf{X}}(t))_{t \in \mathbb{T}}$  from the original metastable Markov process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ . The idea is to define a Galerkin projection based on a division of the state space and to apply it on a transfer operator  $P^t$  (in the following we will in general omit the time  $t$  to make the notation more readable) or the generator  $L$  of  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ . The resulting projected dynamics defined by the projected operators  $QPQ$  or  $QLQ$  can be used to derive the jump intensity and transition rates for the jump process of the Markov state model on the projected state space. We will define in the following all objects necessary for this approach and illustrate it on the case of a full partition of the state space. Afterwards we will also discuss the core set approach that works with a fuzzy partition that is based on the isocommittor functions for a suitable choice of metastable sets, e.g., the neighborhood of an attractor in the double-well example.

### 2.1.1 Galerkin Projection Approach

we will define the basis function for the projection based on a suitable division of the state space  $\mathbb{X}$ . We denote with  $\chi_A$  the characteristic function of a set  $A$ .

**Definition 32.** A family of measurable functions  $\{\Phi_1, \dots, \Phi_m\} \subset L^2(\mu)$  is called a *partition of unity* if it holds that:

1. All functions  $\Phi_k$  are non-negative and linear independent.
2. For the sum over all  $\Phi_k$  it holds that  $\sum_{k=1}^m \Phi_k = \chi_{\mathbb{X}}$  almost everywhere.

A simple example for a partition of unity is the case of a full partition of the state space  $\mathbb{X}$ .

**Example 8.** Let  $\{A_k\}_{k=1, \dots, m}$  be a full partition of the state space  $\mathbb{X}$ , i.e.,  $A_i \cap A_j = \emptyset$  for all  $i \neq j \leq m$  and  $\bigcup_i A_i = \mathbb{X}$ . Then, the characteristic functions  $\{\chi_{A_k}\}_{k=1, \dots, m}$  are a partition of unity, since they are by definition non-negative, linear independent, because the sets  $A_k$  are disjoint, and the sum of the characteristic functions is the characteristic function  $\chi_{\mathbb{X}}$  (the one function on  $\mathbb{X}$ ).

For a given set of basis functions that form a partition of unity for the state space  $\mathbb{X}$  of a Markov process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  with invariant measure  $\mu$  we can define the Galerkin projection that maps operators that act on the space  $L^2(\mu)$  to operators acting on the function space that is defined on the subspace  $D$  spanned by the basis functions.

**Definition 33.** Let  $\{\Phi_1, \dots, \Phi_m\}$  be a partition of unity,  $D := \text{span}\{\Phi_1, \dots, \Phi_m\}$  and  $W$  an  $m \times m$ -matrix with entries  $W(k, j) = \langle \Phi_k, \Phi_j \rangle$ . The matrix  $W$  is invertible, since the  $\Phi_k$  are linear independent. Then, the mapping  $Q : L^2(\mu) \rightarrow D$  with

$$Q\nu = \sum_{k,j=1}^m W^{-1}(k, j) \langle \Phi_k, \nu \rangle \Phi_j$$

is called *Galerkin projection* on the subspace  $D$ .

In the case of a full partition we can write down the Galerkin projection in a slightly simpler way as the scalar product of two different basis functions is zero.

**Example 9.** Let  $\{A_k\}_{k=1, \dots, m}$  be a full partition of the state space  $\mathbb{X}$  and  $\{\Phi_1, \dots, \Phi_m\}$  the respective partition of unity with  $\Phi_k = \chi_{A_k}$  given by the characteristic functions of the partition sets. Then, the matrix  $W$  for the Galerkin projection to the associated subspace  $D$  is given by a diagonal matrix with the entries

$$W(k, k) = \langle \Phi_k, \Phi_k \rangle_{\mu} = \int_{A_k} 1 d\mu(x) = \mu(A_k)$$

and thus  $Q$  can be written as

$$Q\nu = \sum_{k=1}^m \frac{1}{\mu(A_k)} \langle \chi_{A_k}, \nu \rangle_{\mu} \chi_{A_k}$$

### Projection of transfer operators

Now we apply the Galerkin projection  $Q$  to the operators that define the dynamics of a Markov process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  to derive projected versions of these operators that define a Markov process on the reduced state space.

**Definition 34.** Let  $P$  be the transfer operator of a Markov process,  $\Phi_1, \dots, \Phi_n$  a partition of unity and  $Q$  the Galerkin-projection onto the associated subspace  $D$ . The operator  $QPQ : L^2(\mu) \rightarrow D$  is called the *projected transfer operator* onto the subspace  $D$ .

The projected transfer operator can be written down in terms of a matrix representation [13].

**Proposition 2.** [13] Let  $P$  be a transfer operator of a Markov process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ ,  $\Phi_1, \dots, \Phi_n$  a partition of unity and  $QPQ$  the projected transfer operator on the associated subspace. The projected transfer operator  $QPQ$  has a matrix representation

$$P_Q = \mathcal{P}M^{-1},$$

with  $M$  being an  $n \times n$ -matrix with entries

$$M(k, j) = \frac{W(k, j)}{\langle \Phi_k, \chi_{\mathbb{X}} \rangle_{\mu}} = \frac{\langle \Phi_k, \Phi_j \rangle_{\mu}}{\langle \Phi_k, \Phi_{\mathbb{X}} \rangle_{\mu}}$$

and  $\mathcal{P}$  a  $n \times n$ -matrix with entries

$$\mathcal{P}(k, j) = \frac{\langle \Phi_j, P\Phi_k \rangle_{\mu}}{\langle \Phi_k, \chi_{\mathbb{X}} \rangle_{\mu}}.$$

*Proof.* Let  $\{\Psi_1, \dots, \Psi_n\}$  be the basis of the subspace  $D$  with

$$\Psi_k = \frac{\Phi_k}{\langle \Phi_k, \chi_{\mathbb{X}} \rangle_{\mu}}.$$

Then, every function  $u \in D$  can be written as a linear combination of the basis functions

$$u = \sum_{k=1}^n c_k \Psi_k$$

and we can identify  $u$  with the vector of coefficients  $c = (c_k)_{k=1, \dots, n}$ . A matrix representation  $P_Q \in \mathbb{R}^{n \times n}$  of  $QPQ$  needs to map the vector of coefficients  $c$ , such that

$$\sum_{k=1}^n (cP_Q)_k \Psi_k = QPQu$$

holds for all  $u \in D$ . If  $\gamma$  is the vector of coefficients for  $QPQu \in D$  then  $\gamma = cPQ$ . We can write

$$\begin{aligned}
QPQu &= \sum_{k=1}^n c_k QP\Psi_k \\
&= \sum_{k,l,j=1}^n c_k (W^{-1})(l,j) \langle \Phi_j, P\Psi_k \rangle_{\mu} \Phi_l \\
&= \sum_{k,l,j=1}^n c_k (W^{-1})(l,j) \langle \Phi_j, P\Psi_k \rangle_{\mu} \langle \Phi_l, \chi_{\mathbb{X}} \rangle_{\mu} \Psi_l \\
&= \sum_{l=1}^n \gamma_l \Psi_l.
\end{aligned}$$

Then, we have

$$\begin{aligned}
\gamma_l &= \sum_{k,j=1}^n c_k (W^{-1})(l,j) \langle \Phi_j, P\Psi_k \rangle_{\mu} \langle \Phi_l, \chi_{\mathbb{X}} \rangle_{\mu} \\
&= \sum_{k,j=1}^n c_k (W^{-1})(l,j) \frac{\langle \Phi_j, P\Phi_k \rangle_{\mu}}{\langle \Phi_k, \chi_{\mathbb{X}} \rangle_{\mu}} \langle \Phi_l, \chi_{\mathbb{X}} \rangle_{\mu},
\end{aligned}$$

and therefore it holds that

$$P_Q(k,l) = \sum_{j=1}^n (W^{-1})(l,j) \frac{\langle \Phi_j, P\Phi_k \rangle_{\mu}}{\langle \Phi_k, \chi_{\mathbb{X}} \rangle_{\mu}} \langle \Phi_l, \chi_{\mathbb{X}} \rangle_{\mu},$$

which is exactly  $\mathcal{P}M^{-1}(k,l)$ . □

Since we did not explicitly use any properties of the transfer operator itself the same calculation can also be done for the generator  $L$  of the process or any other operator that acts on the space  $L^2(\mu(\mathbb{X}))$ . For the case of a full partition the projected transfer operator can be written down as follows.

**Example 10.** Let  $P$  be a transfer operator of a Markov process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  and  $Q$  be the Galerkin projection for a full partition  $\{A_k\}_{k=1,\dots,m}$  of the state space  $\mathbb{X}$ . Then, the projected transfer operator  $QPQ$  has the matrix representation

$$\begin{aligned}
P_Q(k,l) &= \sum_{j=1}^n (W^{-1})(l,j) \frac{\langle \Phi_j, P\Phi_k \rangle_{\mu}}{\langle \Phi_k, \chi_{\mathbb{X}} \rangle_{\mu}} \langle \Phi_l, \chi_{\mathbb{X}} \rangle_{\mu} \\
&= \frac{1}{\mu(A_l)} \frac{\langle \chi_{A_l}, P\chi_{A_k} \rangle_{\mu}}{\langle \chi_{A_k}, \chi_{\mathbb{X}} \rangle_{\mu}} \langle \chi_{A_l}, \chi_{\mathbb{X}} \rangle_{\mu} \\
&= \frac{\langle \chi_{A_l}, P\chi_{A_k} \rangle_{\mu}}{\langle \chi_{A_k}, \chi_{\mathbb{X}} \rangle_{\mu}}.
\end{aligned}$$

The projected transfer operator  $QPQ$  and projected generator  $QLQ$  inherit many properties of the original transfer operator such as the existence of a dominant eigenvalue 1 with the corresponding stationary distribution being the projected invariant density of the original process [13]. In the full partition case the matrix representation  $P_Q$  of the projected operator  $QPQ$  is a stochastic matrix and induces a discrete time Markov chain between the partition sets, while the matrix representation  $L_Q$  of the operator  $QLQ$  is a rate matrix inducing a continuous-time Markov jump process. We call the process  $(\hat{\mathbf{X}}(t))_{t \in \mathbb{T}}$  that is defined by a projected transfer operator a *Markov state model*. This is, however, only an approximation of the original projected process  $(Q\mathbf{X}(t))_{t \in \mathbb{T}}$ , which in general is no longer Markovian. In general the Galerkin projection and the propagation of the process do not commute, i.e.,  $QP^kQ \neq (QPQ)^k$  and  $Q \exp(tL)Q \neq \exp(tQLQ)$ . The propagation error decays exponentially with the decay rate for the upper bound given by the largest eigenvalue  $\lambda_1$  that is smaller than the dominant eigenvalue  $\lambda_0 = 1$  [69]. It holds that

$$\begin{aligned} E(k) &:= \|QP^kQ - (QPQ)^k\| \\ &\leq \|QP^kQ - \Pi_0\| + \|\Pi_0 - (QPQ)^k\| \\ &\leq \|Q(P^k - \Pi_0)Q\| + \|\Pi_0 - (QPQ)^k\| \leq 2\lambda_1^k \end{aligned}$$

with  $\Pi_0$  denoting the orthogonal projection to the subspace spanned by the invariant measure  $\mu$ , which is the dominant eigenfunction. Since  $\lambda_1$  is usually an eigenvalue that is close to one the error decays quite slowly. The constant of the upper bound can be improved by choosing a sufficiently large lag time and a proper discretization of the state space through the basis functions of the Galerkin projection [69]. One simple strategy for discretizing the state space would be a full partition through a fine grid. This however becomes infeasible in high dimensions and also in low dimensions leads to Markov state models with a high number of states that can be further aggregated. Therefore, we are interested in finding a good set of basis functions for the Galerkin projection  $Q$  that reduces the approximation error of the Markov state model without relying on a large number of partition sets.

### 2.1.2 Metastability

We already mentioned the concept of metastability without giving a formal definition for the meaning of the term. An intuitive characterization is given through the notion of almost invariant sets.

**Definition 35.** Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a Markov process. A set  $A \subset \mathbb{X}$  is called *almost invariant* with respect to time scale  $t \in \mathbb{T}$  if

$$\mathbb{P}(X_t \in A | X_0 \in A) = p(t, A, A) \approx 1.$$

The relation  $p(t, A, A) \approx 1$  can be interpreted as there exists an  $\varepsilon > 0$  such that  $p(t, A, A) \geq 1 - \varepsilon$ .

In the case of multiscale dynamics a set will be considered *metastable* if it is almost invariant for the time scales  $t$  related to the fast dynamics. So, a process will stay with a high probability within a metastable set and the exit time from the set is asymptotically exponential with a small rate parameter [13]. For the full partition Galerkin projection it is useful to consider a partition of the state space into metastable sets.

**Definition 36.** A full partition  $A_1, \dots, A_m$  of a state space  $\mathbb{X}$  is called metastable if every set of the partition is a metastable set. This means that

$$\sum_{k=1}^m \mathbb{P}(X_t \in A_k | X_0 \in A_k) \approx m$$

holds for small time scales  $t$ .

The existence of metastable sets and slow processes is strongly connected to the spectrum of the transfer operator, i.e., the existence of eigenvalues that are close to 1 [13]. The next result from [69] provides a relation between the number of metastable sets of a system and the number of dominant eigenvalues close to 1.

**Proposition 3.** [13] *Let  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  be a reversible Markov process with transfer operator  $P$  fulfilling the following properties:*

1. *The spectral radius of the essential spectrum of  $P$  is smaller than 1.*
2. *For the eigenvalue  $\lambda = 1$  it holds that from  $\eta \in \sigma(P)$  and  $|\eta| = 1$  it follows that  $\eta = 1$ .*

*Then,  $P$  is selfadjoint and the spectrum of  $P$  is of the form:*

$$\sigma(P) \subset [a, b] \cup \{\lambda_m\} \cup \dots \cup \{\lambda_2\} \cup \{1\}$$

*with  $-1 < a \leq b < \lambda_m \leq \dots \leq 1$ . Let  $\nu_1, \dots, \nu_m$  be the respective eigenvectors, normed such that  $\|\nu_k\| = 1$  for all  $k = 1, \dots, m$ . Let  $A_1, \dots, A_m$  be a full partition of the state space  $\mathbb{X}$ ,  $\chi_{A_1}, \dots, \chi_{A_m}$  the respective characteristic functions and  $Q$  the orthogonal projection onto the associated subspace. Then, the sum of the probabilities  $p(A_k, A_k)$  is bounded from above through*

$$p(A_1, A_1) + \dots + p(A_m, A_m) \leq 1 + \lambda_2 + \dots + \lambda_m$$

*and bounded from below through*

$$1 + \kappa_2 \lambda_2 + \dots + \kappa_m + c \lambda_m \leq p(A_1, A_1) + \dots + p(A_m, A_m),$$

*where  $\kappa_j = \|Q\nu_j\|_{L^2(\mu)}^2$  und  $c = a(1 - \kappa_2) \cdots (1 - \kappa_m)$ .*

As we have defined a set to be metastable if it almost invariant, Proposition 3 implies that a partition with more sets than dominant eigenvalues has to include at least one set that is not almost invariant, i.e., the probability to stay within the set is not approximately 1.

The lower bound on the other hand implies that a full partition projection that minimizes the approximation error of the projected eigenfunctions is a metastable partition. While this result answers the question how many metastable sets there are it still remains unclear how to identify them. One approach is through utilizing the eigenfunctions themselves as there is a relation between the sign structure and the metastable sets [71]. Also, the eigenfunctions are almost constant on the metastable sets, so a partition of unity that is based on a metastable partition spans a subspace in which the eigenfunctions can be approximated well [13]. There are however also approaches based on hitting times as the function for the mean first passage time of a properly chosen subset of the state space exhibits also the characteristic of being almost constant on metastable sets [72]. This approach is especially useful in cases where the eigenfunctions can not be calculated, e.g., when the dynamics of the original process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  are not reversible. As the probability mass of the invariant measure also concentrates on the metastable sets, the application of clustering methods to trajectory data is a widely used strategy as well [73].

### 2.1.3 Core Set Approach

Once we have identified the metastable sets, it remains to estimate the parameters for the Markov state model. While for low-dimensional systems it is also possible to directly solve the equation system for the eigenfunctions numerically in general a computational approach is more feasible.

The maximum likelihood estimator for the transition rates between metastable sets  $A_i$  and  $A_j$  of a metastable full partition is given by

$$\kappa_{ij} := \frac{N_{ij}^T}{R_i^T}$$

with  $N_{ij}^T$  counting the transitions from  $A_i$  to  $A_j$  within the time interval  $[0, T]$  and  $R_i^T$  being the total time the process spent within the set  $A_i$  until time  $T$ . For increasing trajectory length  $T \rightarrow \infty$  the estimator  $\kappa_{ij}$  converges to the true value of the transition rate  $\lambda_{ij}$  of the projected process. Of course we are limited by the finiteness of the trajectory data that we can generate, but besides that there is also another source of error that is related to how we define what we consider a transition between metastable sets.

#### Recrossing problem

As already was mentioned before through the projection of the process we lose in general the Markov property. This can be illustrated by the example of the one-dimensional double-well potential. Let us consider a partition of the state space into two sets  $A_1 = (-\infty, 0]$ ,  $A_2 = (0, \infty)$  with the barrier separating the sets being the local maximum of the potential. When a transition from  $A_1$  to  $A_2$  happens then the process will still be close to the local maximum directly after the transition and have a significantly higher probability to recross the barrier between the sets compared to the average probability

for an arbitrary position within  $A_2$ . In this regard, the projected process has a memory effect and is no longer Markovian, as it does matter whether the set  $A_2$  was just entered or the process has remained there already for a longer time. If we count every crossing of the barrier between the two sets as a transition, then we will overestimate the transition rate with the maximum likelihood estimator. One solution that reduces this error is to introduce a lag time  $\tau > \Delta t$  to define that a transition from  $A_i$  to  $A_j$  after lag time  $\tau$  occurs when  $\mathbf{X}(t) \in A_1$  and  $\mathbf{X}(t + \tau) \in A_2$ . This leads to only a part of the trajectory data being used and some of the recrossings during the transition between the two metastable sets being omitted. For a sufficiently large lag time  $\tau$  this procedure reduces the error of the estimator significantly [13].

### Milestoning process

An alternative approach to the problem of recrossings utilizes a discretization approach that is not based on a full partition of the state space. In the core set approach for each set  $A_k$  of the metastable full partition a core set  $C_k$  is chosen that focuses on the local neighborhood of an attractor and does not include parts of the transition region. The core sets can be chosen to be metastable sets themselves but because of the fast mixing within the metastable sets also a proper subset can be sufficient for the purpose. We define the milestoning process  $(\tilde{\mathbf{X}}(t))_{t \in \mathbb{T}}$  for the estimation of the transition rates of the original process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  between the core sets utilizing memory about the past of the trajectory [74]. The process  $(\tilde{\mathbf{X}}(t))_{t \in \mathbb{T}}$  takes as value the index of the last core set that was visited by  $(\mathbf{X}(t))_{t \in \mathbb{T}}$ , i.e

$$\tilde{\mathbf{X}}(t) = i \Leftrightarrow \mathbf{X}_{\sigma(t)} \in C_i$$

for the stopping time

$$\sigma(t) := \sup_{s \leq t} \left\{ s : \mathbf{X}(s) \in \bigcup_{k=1}^m C_k \right\}$$

which tracks the last time that the process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  was within one of the  $m$  core sets. As the definition of the milestoning process  $(\tilde{\mathbf{X}}(t))_{t \in \mathbb{T}}$  involves a stopping time with memory, it is clear that it is not Markovian. An alternative construction of the milestoning process would be to introduce an extended state space  $\mathbb{Y} := \mathbb{X} \times \{1, \dots, m\}$  and define a process  $(\mathbf{Y}(t))_{t \in \mathbb{T}}$  where the first component is given by the original process  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  and the second component is updated to a new index whenever  $(\mathbf{X}(t))_{t \in \mathbb{T}}$  enters a core set with a different index than the current value of the second component. This combined process would be Markovian as all information that would lead to a dependence on the trajectory is incorporated in the present state and the marginal process of the second component would coincide with the milestoning process  $(\tilde{\mathbf{X}}(t))_{t \in \mathbb{T}}$ . The projection onto the marginal process of the second component would however again lead to a loss of the Markov property. The projection approach using a full partition has the recrossing problem where we know that right after entering a new partition set the probability for leaving is higher than average. We have a similar situation for the milestoning process  $(\tilde{\mathbf{X}}(t))_{t \in \mathbb{T}}$ . At the time when a new value is taken we know that the original process is

within one of the core sets and not near the transition region and that the probability for entering another core set is lower than average [13]. However, for the estimation of the transition rates between the core sets this memory effect for the milestoning process is negligible.

### Core set MSM

Thus, we can consider a set of  $m$  core sets  $C_1, \dots, C_m$  instead of a metastable full partition  $A_1, \dots, A_m$  for the construction of an MSM  $(\hat{\mathbf{X}}(t))_{t \in \mathbb{T}}$  and count the transitions between the core sets for the maximum likelihood estimator  $\kappa_{ij}$  and define a jump process according to the calculated transition rates. This procedure is also associated with the application of a Galerkin projection, but with a set of committor functions as basis for the projection operator. The basis functions associated with the core set MSM are the isocommittor functions  $q_1, \dots, q_m$  defined by

$$q_k := q_{C_k, \bar{C}_k}. \quad (2.1)$$

The committor function for entering  $C_k$  before the set  $\bar{C}_k$  which is defined as the union of all other core sets

$$\bar{C}_k := \bigcup_{\substack{l=1 \\ l \neq k}}^m C_l.$$

The projection error of the core set MSM associated with the isocommittor projection is smaller than that of a full partition for the right choice of core sets [69]. The jump process defined by the core set MSM  $(\hat{\mathbf{X}}(t))_{t \in \mathbb{T}}$  has then the same transition rates as the milestoning process  $(\tilde{\mathbf{X}}(t))_{t \in \mathbb{T}}$  [75].

## 2.2 Stochastic Metapopulation Model

As we motivated the Markov state model construction with the diffusion in a double-well potential, the metastability of the guiding Example 6 shall be our motivation for this section. The diffusion process from Example 6 exhibits metastable behavior: An agent remains for a comparatively long period of time within one well of the potential  $U$  before it eventually jumps to the other well. The metastability of the mobility dynamics also induces a metastability in the adoption dynamics. Until the critical transition event happens the spreading of status 2 happens only in one of the two wells. The Markov state modeling approach lets us derive a reduced model for the mobility of the process, but also the adoption dynamics has to be adapted to the reduced state space such that the MSM can be related to the projected ABM. The resulting model is a metapopulation model [76], where the total population is divided into subpopulations (which in our guiding example represent the metastable sets) with rare spatial transitions between the subpopulations and internal adoption dynamics within the subpopulations. The diffusive characteristics of the mobility process gets reduced to that of a pure jump process through the projection and thus the stochastic metapopulation model (SMM) is a pure jump process.

In the microscopic dynamics of ABM, not all agents within a metastable set can interact with each other. The pairwise second order adoptions require that the distance in physical space between two agents is sufficiently small, but in the reduced state space, where we cannot distinguish between different positions within a metastable set, all individuals within a subpopulation are connected. The mesoscopic rate constants for the adoption rate functions of the coarse-grained model must then be adjusted relative to the microscale rate constants so that the adoption dynamics of the metapopulation model reflect the adoption dynamics of the projected ABM.

In this section, we will first provide a formal definition of the SMM and then establish a relation to the projected ABM by applying the MSM framework. By applying a Galerkin projection with an appropriate set of basis functions, we can establish an explicit relation between the microscopic and mesoscopic rate constants of the adoption rate functions. Overall, this section is structured similarly to [16], and the notation and proofs are mostly done in the same way.

### 2.2.1 SMM Definition

We consider a population with  $n_a$  members (agents) that is divided into  $m$  subpopulations and the finite status space  $\mathbb{S} = \{1, \dots, n_s\}$ . The system state of the stochastic metapopulation model at time  $t$  is denoted by an  $n_s \times m$  matrix  $\mathbf{N}(t) = (\mathbf{N}_i^{(k)}(t))_{i=1, \dots, n_s, k=1, \dots, m}$ , where  $\mathbf{N}_i^{(k)}(t)$  refers to the number of members (agents) of the subpopulation  $k$  in status  $i$  at time  $t$ . The set of all possible system states for a system with  $n_a$  agents, is denoted by  $\mathbb{M}_{n_a}$ ,

$$\mathbb{M}_{n_a} := \left\{ N = (N_i^{(k)})_{i=1, \dots, n_s, k=1, \dots, m} \in \mathbb{N}_0^{n_s, m} : \sum_{i=1}^{n_s} \sum_{k=1}^m N_i^{(k)} = n_a \right\}. \quad (2.2)$$

We distinguish between two types of jump events: spatial transitions between the subpopulations and status adoptions within a subpopulation. A member of status  $i$  in subpopulation  $k$  transitioning to subpopulation  $l$  at time  $t$  is associated with the state change vector  $-E_i^{(k)} + E_i^{(l)}$ , where  $E_i^{(k)}$  is an  $n_s \times m$  matrix with all entries zero except for the entry at index  $(k, i)$  being one. Similarly, an adoption event from status  $i$  to status  $j$  in subpopulation  $k$  leads to jump event associated with the state change vector  $-E_i^{(k)} + E_j^{(k)}$ .

Let  $P(N, t) := \mathbb{P}(\mathbf{N}(t) = N)$  denote the probability to find the system in state  $N \in \mathbb{M}_{n_a}$  at time  $t$ . Then, the evolution of the metapopulation model is given by the master equation

$$\frac{d}{dt} P(N, t) = \mathcal{L}P(N, t) + \mathcal{G}P(N, t) \quad (2.3)$$

for operators  $\mathcal{L}, \mathcal{G}$  given by

$$\mathcal{L}P(M) := \sum_{N \in \mathbb{M}_{n_a}} \hat{\mathcal{L}}_{NM} \cdot P(N), \quad \mathcal{G}P(M) := \sum_{N \in \mathbb{M}_{n_a}} \hat{\mathcal{G}}_{NM} \cdot P(N), \quad (2.4)$$

where  $\mathcal{L}$  is a rate matrix, i.e.,  $\hat{\mathcal{L}}_{NM}$  for  $M \neq N$  is the rate to go from  $N$  to  $M$  by a spatial transition event between the subpopulations and  $\hat{\mathcal{L}}_{NN} := -\sum_M \hat{\mathcal{L}}_{NM}$ . Analogously,  $\mathcal{G}$  is a rate matrix for changes through adoption events, where the entries are specified by adoption rate functions  $\hat{f}_{ij}^{(k)} : \mathbb{M}_{n_a} \rightarrow [0, \infty)$  for status changes from  $i$  to  $j$  in subpopulation  $k$ . In the following we specify the shape of  $\hat{\mathcal{L}}$  and  $\hat{\mathcal{G}}$  and show their connection to the ABM operators  $L$  and  $G$  of Equation (1.12) by means of Galerkin projection methods.

## 2.2.2 Galerkin Projection Approach for ABMs

Let us assume that the mobility process of an ABM exhibits metastable dynamics and that we can find a partition of the state space  $\mathbb{X}$  into  $m$  metastable sets  $A_1, \dots, A_m$ . We then make use of the approach given in [16,77] to define the basis functions for a Galerkin projection  $Q$  to construct a Markov state model with respect to the movements of the agents. We apply the projection operator  $Q$  to the generator  $L + G$  defined in (1.12) and derive a projected generator, see also [78]. In particular, we will study how this projection affects the agents' mobility process and the adoption rate functions, especially those for first-order and second-order adoptions defined in Equations (1.6) and (1.7). For the derivation of the analytical results, we will focus on the case of a full partition of the state space as is described in Section 2.1.1.

### Basis functions of the projection

For any  $N = (N_i^{(k)}) \in \mathbb{M}_{n_a}$  we define the indicator ansatz functions

$$\Phi_N(X, S) := \prod_{k=1}^m \prod_{i=1}^{n_s} \phi_{N_i^{(k)}}(X, S) \quad (2.5)$$

with

$$\phi_{N_i^{(k)}}(X, S) := \delta_{N_i^{(k)}} \left( \sum_{\alpha=1}^{n_a} \chi_{A_k}(x_\alpha) \delta_i(s_\alpha) \right), \quad (2.6)$$

where  $\delta_i$  denotes the Kronecker delta for status  $i$  and  $\chi_{A_k}$  the characteristic function of set  $A_k$ . The function  $\Phi_N(X, S)$  has the value 1 whenever there are for each  $i, k$  exactly  $N_i^{(k)}$  agents  $\alpha$  with position  $x_\alpha \in A_k$  and status  $s_\alpha = i$ , otherwise it is zero. These ansatz functions are by definition non-negative and associate each ABM system state uniquely with a metapopulation system state  $N \in \mathbb{M}_{n_a}$ , i.e.,  $\sum_{N \in \mathbb{M}_{n_a}} \Phi_N(X, S) = 1$  for all  $(X, S) \in \mathbb{Y}$ . Thus the functions  $\Phi_N$  form a partition of unity.

Next, we define the inner product of two functions  $f, g : \mathbb{Y} \rightarrow \mathbb{R}$  as

$$\langle f, g \rangle := \frac{1}{(\mu(\mathbb{X})n_s)^{n_a}} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} f(X, S)g(X, S) dX, \quad (2.7)$$

where  $\mu$  denotes the Lebesgue measure on  $\mathbb{X}$ . For the indicator ansatz functions  $\Phi_N$  defined above we observe that  $\langle \Phi_M, \Phi_N \rangle = 0$  holds for  $M \neq N$ , while for  $M = N$  we have  $\langle \Phi_M, \Phi_N \rangle = \langle \Phi_N, \Phi_N \rangle = \langle \Phi_N, \mathbb{1} \rangle$ , where  $\mathbb{1}$  denotes the constant 1-function on  $\mathbb{Y}$ .

### Definition of the projection operator

Now, after defining the basis functions and the inner product for functions on the system state space, we can define the full-partition projection  $Q : L^2(\mathbb{Y}) \rightarrow D$  to the ansatz

space  $D = \text{span}\{\Phi_N, N \in \mathbb{M}_{n_a}\}$  given by [13]

$$Qv = \sum_{N \in \mathbb{M}_{n_a}} \frac{\langle \Phi_N, v \rangle}{\langle \Phi_N, \mathbb{1} \rangle} \Phi_N. \quad (2.8)$$

Given any linear operator  $H : L^2(\mathbb{Y}) \rightarrow L^2(\mathbb{Y})$ , a Galerkin projection with  $Q$  yields the *projected operator*  $QHQ : L^2(\mathbb{Y}) \rightarrow D$ . The goal is now to find the matrix representations  $\hat{\mathcal{L}} = (\hat{\mathcal{L}}_{NM})_{N, M \in \mathbb{M}_{n_a}}$  and  $\hat{\mathcal{G}} = (\hat{\mathcal{G}}_{NM})_{N, M \in \mathbb{M}_{n_a}}$  (see Equation (2.4)) of the projected operators  $QLQ$  and  $QGQ$  for the operators  $L$  and  $G$  defined in (1.10) and (1.11), respectively. The jump process of the stochastic metapopulation model that is defined through the matrix representations is then the MSM approximation for the projected ABM process.

The following corollaries from [16] will be needed to show our main results of this section. Again, we use the notation  $e_\alpha$  for the  $\alpha$ -th unit vector of  $\mathbb{R}^{n_a}$ , while  $E_i^{(k)}$  denotes an  $n_s \times m$  matrix with all entries zero except the entry at index  $(k, i)$  which is one.

**Corollary 1.** [16] *For any  $N \in \mathbb{M}_{n_a}$  and given  $i, j \in \mathbb{S}$ , it holds that*

$$\Phi_N(X, S + ie_\alpha - je_\alpha) = \Phi_{N + E_j^{(k_\alpha)} - E_i^{(k_\alpha)}}(X, S)$$

for each  $\alpha \in \{1, \dots, n_a\}$  with  $s_\alpha = j$  and  $N_i^{(k_\alpha)} > 0$ .

The condition  $N_i^{(k_\alpha)} > 0$  in Corollary 1 is necessary to guarantee that  $N + E_j^{(k_\alpha)} - E_i^{(k_\alpha)} \in \mathbb{M}_{n_a}$  holds, so that  $\Phi_{N + E_j^{(k_\alpha)} - E_i^{(k_\alpha)}}$  is actually defined. In order to simplify the notation in all the following calculations, we extend the definition of  $\Phi_N$  and set  $\Phi_N(X, S) := 0$  for  $N \notin \mathbb{M}_{n_a}$ . With this definition, Corollary 1 also works for  $N_i^{(k)} = 0$  because both sides of the equation become zero.

**Proof.** Choose  $\ell \in \{1, \dots, n_s\}$  and  $k \in \{1, \dots, m\}$ . For  $i \neq j$  we consider

$$\begin{aligned} & \delta_j(s_\alpha) \phi_{N_\ell^{(k)}}(X, S + ie_\alpha - je_\alpha) \\ \stackrel{(2.6)}{=} & \delta_j(s_\alpha) \cdot \delta_{N_\ell^{(k)}} \left( \sum_{\beta=1}^{n_a} \chi_{A_k}(x_\beta) \begin{cases} \delta_\ell(s_\beta) & \text{if } \beta \neq \alpha \\ \delta_\ell(i) & \text{if } \beta = \alpha \end{cases} \right) \\ = & \delta_j(s_\alpha) \cdot \delta_{N_\ell^{(k)}} \left( \begin{cases} \sum_{\beta=1}^{n_a} \chi_{A_k}(x_\beta) \delta_\ell(s_\beta) & \text{if } x_\alpha \notin A_k \\ \sum_{\beta=1}^{n_a} \chi_{A_k}(x_\beta) \delta_\ell(s_\beta) + \delta_\ell(i) - \delta_\ell(j) & \text{if } x_\alpha \in A_k \end{cases} \right). \end{aligned}$$

Thus, using definition (2.6), we get for  $k$  with  $x_\alpha \notin A_k$ :

$$\delta_j(s_\alpha) \phi_{N_\ell^{(k)}}(X, S + ie_\alpha - je_\alpha) = \delta_j(s_\alpha) \phi_{N_\ell^{(k)}}(X, S)$$

for any  $\ell$ . For  $k$  such that  $x_\alpha \in A_k$ , on the other hand, we distinguish between the following cases. For  $\ell \neq i$  and  $\ell \neq j$ , it holds

$$\delta_j(s_\alpha) \phi_{N_\ell^{(k)}}(X, S + ie_\alpha - je_\alpha) = \delta_j(s_\alpha) \phi_{N_\ell^{(k)}}(X, S),$$

for  $\ell = j$ , we calculate

$$\begin{aligned}
\delta_j(s_\alpha)\phi_{N_j^{(k)}}(X, S + ie_\alpha - je_\alpha) &= \delta_j(s_\alpha)\delta_{N_j^{(k)}}\left(\sum_{\beta=1}^{n_a}\chi_{A_k}(x_\beta)\delta_j(s_\beta) - 1\right) \\
&= \delta_j(s_\alpha)\delta_{N_j^{(k)+1}}\left(\sum_{\beta=1}^{n_a}\chi_{A_k}(x_\beta)\delta_j(s_\beta)\right) \\
&= \delta_j(s_\alpha)\phi_{N_j^{(k)+1}}(X, S)
\end{aligned}$$

and for  $\ell = i$ , we analogously get

$$\delta_j(s_\alpha)\phi_{N_i^{(k)}}(X, S + ie_\alpha - je_\alpha) = \delta_j(s_\alpha)\phi_{N_i^{(k)-1}}(X, S).$$

By definition of  $E_i^{(k)}$  the matrix  $N + E_j^{(k)} - E_i^{(k)}$  is the state where all numbers  $N_\ell^{(k)}$  stay the same except  $N_i^{(k)}$ , which is replaced by  $N_i^{(k)} - 1$ , and  $N_j^{(k)}$ , which is replaced by  $N_j^{(k)} + 1$ . Let  $k_\alpha$  denote the index of the set  $A_k$  for which  $x_\alpha \in A_k$ . Then, combining the above calculations and using the definition  $\Phi_N = \prod_{k=1}^m \prod_{i=1}^{n_s} \phi_{N_i^{(k)}}$  of  $\Phi_N$  given in (2.5), we obtain

$$\Phi_N(X, S + ie_\alpha - je_\alpha) = \Phi_{N+E_j^{(k_\alpha)}-E_i^{(k_\alpha)}}(X, S)$$

for each  $\alpha = 1, \dots, n_a$  with  $\delta_j(s_\alpha) = 1$ . □

For the basis functions of metapopulation system states that differ by only one adoption event, we also need the following relationship.

**Corollary 2.** [16] For  $N \in \mathbb{M}_{n_a}$  with  $N_i^{(k)} > 0$  and  $M = N + E_j^{(k)} - E_i^{(k)}$  it holds that

$$\frac{\langle \Phi_M, \Phi_M \rangle}{\langle \Phi_N, \Phi_N \rangle} = \frac{N_i^{(k)}}{N_j^{(k)} + 1}. \quad (2.9)$$

**Proof.** Using basic combinatorics, we obtain that

$$\begin{aligned}
\langle \Phi_N, \Phi_N \rangle &= \frac{1}{(\mu(\mathbb{X})n_s)^{n_a}} \int_{\mathbb{X}^{n_a}} \sum_{S \in \mathbb{S}^{n_a}} \Phi_N(X, S) dX \\
&= \frac{n_a!}{\prod_{\kappa, \ell} N_\ell^{(\kappa)}!} \prod_{\kappa, \ell} \left( \frac{\mu(A_\kappa)}{n_s \mu(\mathbb{X})} \right)^{N_\ell^{(\kappa)}}. \quad (2.10)
\end{aligned}$$

This results from the multinomial distribution of  $n_a$  agents into boxes  $(\kappa, \ell)$ ,  $\kappa = 1, \dots, m$ ,  $\ell = 1, \dots, n_s$  with  $N_\ell^{(\kappa)}$  agents each and the box probabilities  $p_{\kappa, \ell} := \frac{\mu(A_\kappa)}{n_s \mu(\mathbb{X})}$ . Then, for  $M = N + E_j^{(k)} - E_i^{(k)}$  by using Equation (2.10) we directly obtain (2.9). □

Finally, given a linear operator  $H : L^2(\mathbb{Y}) \rightarrow L^2(\mathbb{Y})$ , we derive the matrix representation of the projected operator  $QHQ$  for the considered case of a full-partition projection.

**Corollary 3.** [16] Given a linear operator  $H : L^2(\mathbb{Y}) \rightarrow L^2(\mathbb{Y})$ , the Galerkin projection  $QHQ$  with  $Q$  defined in (2.8) has the matrix representation  $\hat{\mathcal{H}}$ , where

$$\hat{\mathcal{H}}_{NM} = \frac{\langle \Phi_M, H\Phi_N \rangle}{\langle \Phi_N, \mathbb{1} \rangle}. \quad (2.11)$$

The calculation for the corollary is analogous to the proof of Proposition 2 with the chosen basis functions from Equation (2.6).

**Remark 1.** In the case of more general basis functions  $\Phi_N$  that form a partition of unity but are not necessarily indicator functions, e.g., a basis of committor functions in case of the core set approach, the result from Corollary 3 can be extended by applying Proposition 2 to the more general basis. The matrix representation of the projected generator  $QHQ$  with  $Q$  given by

$$Qv = \sum_{M,N \in \mathbb{M}_{n_a}} (W^{-1})_{MN} \langle \Phi_M, v \rangle \Phi_N, \quad W_{MN} := \langle \Phi_M, \Phi_N \rangle \quad (2.12)$$

has the form  $\hat{\mathcal{H}}\hat{W}^{-1}$  with  $\hat{\mathcal{H}}$  given in (2.11) and

$$\hat{W}_{MN} := \frac{\langle \Phi_N, \Phi_M \rangle}{\langle \Phi_M, \mathbb{1} \rangle}.$$

In the following, we will use these corollaries for proving the main results (Theorems 37-39) for the explicit relation between the mesoscale and microscale rate constants for the spatial transitions and the first and second order adoptions.

### 2.2.3 Projection of the Spatial Dynamics

We consider the generator  $L$  of the mobility process of the ABM system and define

$$\lambda_i^{(kl)} := \frac{\langle \chi_{A_l}, L_i \chi_{A_k} \rangle_{\mathbb{X}}}{\langle \chi_{A_k}, \mathbb{1} \rangle_{\mathbb{X}}} = \frac{\int_{\mathbb{X}} \chi_{A_l}(x) (L_i \chi_{A_k})(x) dx}{\int_{\mathbb{X}} \chi_{A_k}(x) dx} \quad (2.13)$$

where  $\langle \cdot, \cdot \rangle_{\mathbb{X}}$  refers to the standard scalar product for functions in  $L^2(\mathbb{X})$  and  $\mathbb{1}$  denotes the constant 1-function on  $\mathbb{X}$  [16].

**Theorem 37.** The matrix representation of the projected generator  $QLQ$  is given by  $\hat{\mathcal{L}}$  with

$$\hat{\mathcal{L}}_{NM} = \begin{cases} \lambda_i^{(kl)} N_i^{(k)}, & \text{if } M = N + E_i^{(l)} - E_i^{(k)}, k \neq l, \\ -\sum_{i=1}^{n_s} \sum_{k,l=1;l \neq k}^m \lambda_i^{(kl)} N_i^{(k)}, & \text{if } M = N, \\ 0, & \text{otherwise.} \end{cases}$$

*Proof.* We first observe that for fixed  $\alpha \in \{1, \dots, n_a\}$  and  $i \in \mathbb{S}$  it holds

$$\delta_i(s_\alpha) \Phi_N(X, S) = \sum_{k=1}^m \delta_k(x_\alpha) \delta_i(s_\alpha) \Phi_{N-E_i^{(k)}}(X^{-\alpha}, S^{-\alpha}), \quad (2.14)$$

where  $S^{-\alpha} \in \mathbb{S}^{n_a-1}$  denotes the vector resulting from  $S \in \mathbb{S}^{n_a}$  by skipping the entry  $s_\alpha$ , and  $X^{-\alpha} \in \mathbb{X}^{n_a}$  results from  $X \in \mathbb{X}^{n_a}$  by skipping the entry  $x_\alpha$ .<sup>1</sup>

Set  $\mu_0 := \frac{1}{(\mu(\mathbb{X})n_s)^{n_a}}$ . Using the definition of  $L$  given in (1.10), we calculate

$$\begin{aligned}
& \langle \Phi_M, L\Phi_N \rangle \\
&= \mu_0 \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \sum_{\alpha=1}^{n_a} \sum_{i=1}^{n_s} \delta_i(s_\alpha) L_i^{(\alpha)} \Phi_N(X, S) dX \\
&= \mu_0 \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \sum_{\alpha=1}^{n_a} \sum_{i=1}^{n_s} \sum_{l=1}^m \chi_{A_l}(x_\alpha) \delta_i(s_\alpha) L_i^{(\alpha)} \Phi_N(X, S) dX \\
(2.14) \quad &= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i=1}^{n_s} \sum_{k,l=1}^m \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_{M-E_i^{(l)}}(X^{-\alpha}, S^{-\alpha}) \chi_{A_l}(x_\alpha) \delta_i(s_\alpha) (L_i \chi_{A_k})(x_\alpha) \Phi_{N-E_i^{(k)}}(X^{-\alpha}, S^{-\alpha}) dX \\
&= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i=1}^{n_s} \sum_{k,l=1}^m \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a-1}} \Phi_{M-E_i^{(l)}}(X^{-\alpha}, S^{-\alpha}) \delta_i(s_\alpha) \int_{\mathbb{X}} \chi_{A_l}(x_\alpha) (L_i \chi_{A_k})(x_\alpha) dx_\alpha \Phi_{N-E_i^{(k)}}(X^{-\alpha}, S^{-\alpha}) dX^{-\alpha} \\
(2.13) \quad &= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i=1}^{n_s} \sum_{k,l=1}^m \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a-1}} \Phi_{M-E_i^{(l)}}(X^{-\alpha}, S^{-\alpha}) \delta_i(s_\alpha) \lambda_i^{(kl)} \int_{\mathbb{X}} \chi_{A_k}(x_\alpha) dx_\alpha \Phi_{N-E_i^{(k)}}(X^{-\alpha}, S^{-\alpha}) dX^{-\alpha} \\
&= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i=1}^{n_s} \sum_{k,l=1}^m \lambda_i^{(kl)} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_{M-E_i^{(l)}+E_i^{(k)}}(X, S) \delta_i(s_\alpha) \chi_{A_k}(x_\alpha) \Phi_N(X, S) dX \\
&= \mu_0 \sum_{i=1}^{n_s} \sum_{k,l=1}^m \lambda_i^{(kl)} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_{M-E_i^{(l)}+E_i^{(k)}}(X, S) \sum_{\alpha=1}^{n_a} \delta_i(s_\alpha) \chi_{A_k}(x_\alpha) \Phi_N(X, S) dX \\
(*) \quad &= \mu_0 \sum_{i=1}^{n_s} \sum_{k,l=1}^m \lambda_i^{(kl)} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_{M-E_i^{(l)}+E_i^{(k)}}(X, S) N_i^{(k)} \Phi_N(X, S) dX \\
&= \sum_{i=1}^{n_s} \sum_{k,l=1}^m \lambda_i^{(kl)} N_i^{(k)} \langle \Phi_{M-E_i^{(l)}+E_i^{(k)}}, \Phi_N \rangle,
\end{aligned}$$

where (\*) results from the fact that it holds  $\sum_{\alpha=1}^{n_a} \delta_i(s_\alpha) \chi_{A_k}(x_\alpha) = N_i^{(k)}$  for all  $(X, S)$  with  $\Phi_N(X, S) \neq 0$ . Assume that it holds  $M = N + E_i^{(l)} - E_i^{(k)}$  for certain  $k, l, i, k \neq l$ . Then, all summands are zero except one summand and we obtain

$$\langle \Phi_M, L\Phi_N \rangle = \lambda_i^{(kl)} N_i^{(k)} \langle \Phi_N, \Phi_N \rangle.$$

For the case  $M = N$  we need  $k = l$  such that  $\Phi_{M-E_i^{(l)}+E_i^{(k)}} = \Phi_M = \Phi_N$ , and obtain

$$\langle \Phi_M, L\Phi_N \rangle = \sum_{i=1}^{n_s} \sum_{k=1}^m \lambda_i^{(kk)} N_i^{(k)} \langle \Phi_N, \Phi_N \rangle.$$

<sup>1</sup>Again, we use the extended definition  $\Phi_N := 0$  for  $N \notin \mathbb{M}_{n_a}$ , such that  $\Phi_{N-E_i^{(k)}}$  is also defined in case of  $N_i^{(k)} = 0$ .

For other combinations of  $M, N$  the overall sum is zero. In total, we get

$$\begin{aligned}
& \langle \Phi_M, L\Phi_N \rangle \\
= & \begin{cases} N_i^{(k)} \lambda_i^{(kl)} \langle \Phi_N, \Phi_N \rangle, & \text{if } M = N + E_i^{(l)} - E_i^{(k)}, k \neq l \\ \sum_{i=1}^{n_s} \sum_{k=1}^m N_i^{(k)} \lambda_i^{(kk)} \langle \Phi_N, \Phi_N \rangle, & \text{if } M = N, \\ 0, & \text{otherwise,} \end{cases} \\
= & \begin{cases} N_i^{(k)} \lambda_i^{(kl)} \langle \Phi_N, \Phi_N \rangle, & \text{if } M = N + E_i^{(l)} - E_i^{(k)}, k \neq l \\ - \sum_{i=1}^{n_s} \sum_{k,l=1l \neq k}^m N_i^{(k)} \lambda_i^{(kl)} \langle \Phi_N, \Phi_N \rangle, & \text{if } M = N, \\ 0, & \text{otherwise,} \end{cases}
\end{aligned}$$

where for the case  $M = N$  we used  $\lambda_i^{(kk)} = - \sum_{l \neq k} \lambda_i^{(kl)}$ . By means of Corollary 3 we can have to divide by  $\langle \Phi_N, \mathbb{1} \rangle = \langle \Phi_N, \Phi_N \rangle$  and obtain

$$\hat{\mathcal{L}}_{NM} = \begin{cases} N_i^{(k)} \lambda_i^{(kl)}, & \text{if } M = N + E_i^{(l)} - E_i^{(k)}, k \neq l \\ - \sum_{i=1}^{n_s} \sum_{k,l=1l \neq k}^m N_i^{(k)} \lambda_i^{(kl)}, & \text{if } M = N, \\ 0, & \text{otherwise} \end{cases}$$

for the entries of  $\hat{\mathcal{L}}$ . □

## 2.2.4 Projection of the Adoption Dynamics

Now we derive the matrix representations of the projected generator  $QGQ$  for the adoption dynamics. We consider the two special cases of first-order and second-order adoptions separately.

### First-order adoptions

For first-order status changes with adoption rate functions given in (1.6), we define the conditional expectation of  $\gamma_{ij}(x)$  given that  $x \in A_k$ :

$$\gamma_{ij}^{(k)} := \frac{\langle \gamma_{ij}, \chi_{A_k} \rangle_{\mathbb{X}}}{\langle \chi_{A_k}, \mathbb{1} \rangle_{\mathbb{X}}} = \frac{\int_{\mathbb{X}} \gamma_{ij}(x) \chi_{A_k}(x) dx}{\int_{\mathbb{X}} \chi_{A_k}(x) dx}. \quad (2.15)$$

Then, we obtain the following result.

**Theorem 38.** *For first-order adoptions with an ABM rate function  $f_{ij}^{(\alpha)}$  of the form (1.6), the projected generator  $QGQ$  has the matrix representation  $\hat{\mathcal{G}}$  with*

$$\hat{\mathcal{G}}_{NM} = \begin{cases} \hat{f}_{ij}^{(k)}(N), & \text{if } M = N + E_j^{(k)} - E_i^{(k)}, i \neq j, \\ -\sum_{i,j=1}^{n_s} \sum_{k=1}^m \hat{f}_{ij}^{(k)}(N), & \text{if } M = N, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$\hat{f}_{ij}^{(k)}(N) := \gamma_{ij}^{(k)} N_i^{(k)}.$$

*Proof.* Based on the Corollary 1, we see that the action of the ABM generator  $G$  on an individual indicator ansatz function can be written as

$$\begin{aligned} G\Phi_N(X, S) &= \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} \left( -f_{ij}^{(\alpha)}(X, S) \Phi_N(X, S) + f_{ij}^{(\alpha)}(X, S + ie_\alpha - je_\alpha) \Phi_{N+E_j^{(k\alpha)}-E_i^{(k\alpha)}}(X, S) \right), \end{aligned}$$

with the consequence that

$$\mathcal{G}_{MN} := \langle \Phi_M, G\Phi_N \rangle = -\mathcal{G}_{1,M,N} + \mathcal{G}_{2,M,N}, \quad (2.16)$$

where

$$\begin{aligned} \mathcal{G}_{1,M,N} &:= \left\langle \Phi_M, \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)} \Phi_N \right\rangle \\ &= \mu_0 \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)}(X, S) \Phi_N(X, S) dX, \\ \mathcal{G}_{2,M,N} &:= \left\langle \Phi_M, \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)}(X, S + ie_\alpha - je_\alpha) \Phi_{N+E_j^{(k\alpha)}-E_i^{(k\alpha)}} \right\rangle \\ &= \mu_0 \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)}(X, S + ie_\alpha - je_\alpha) \Phi_{N+E_j^{(k\alpha)}-E_i^{(k\alpha)}}(X, S) dX \end{aligned}$$

and again  $\mu_0 := \frac{1}{(\mu(\mathbb{X})n_s)^{n_a}}$ . We compute

$$\begin{aligned}
& \mathcal{G}_{1,M,N} \\
&= \mu_0 \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)}(X, S) \Phi_N(X, S) dX \\
&= \mu_0 \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} \delta_i(s_\alpha) \gamma_{ij}(x_\alpha) \Phi_N(X, S) dX \\
&= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i,j=1}^{n_s} \sum_{k=1}^m \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \delta_i(s_\alpha) \chi_{A_k}(x_\alpha) \gamma_{ij}(x_\alpha) \Phi_N(X, S) dX \\
&= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i,j=1}^{n_s} \sum_{k=1}^m \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a-1}} \Phi_{M-E_i^{(k)}}(X^{-\alpha}, S^{-\alpha}) \delta_i(s_\alpha) \int_X \chi_{A_k}(x_\alpha) \gamma_{ij}(x_\alpha) dx_\alpha \Phi_{N-E_i^{(k)}}(X^{-\alpha}, S^{-\alpha}) dX^{-\alpha} \\
&\stackrel{(2.15)}{=} \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i,j=1}^{n_s} \sum_{k=1}^m \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a-1}} \Phi_{M-E_i^{(k)}}(X^{-\alpha}, S^{-\alpha}) \delta_i(s_\alpha) \gamma_{ij}^{(k)} \int_X \chi_{A_k}(x_\alpha) dx_\alpha \Phi_{N-E_i^{(k)}}(X^{-\alpha}, S^{-\alpha}) dX^{-\alpha} \\
&= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i,j=1}^{n_s} \sum_{k=1}^m \gamma_{ij}^{(k)} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \delta_i(s_\alpha) \chi_{A_k}(x_\alpha) \Phi_N(X, S) dX \\
&= \mu_0 \sum_{i,j=1}^{n_s} \sum_{k=1}^m \gamma_{ij}^{(k)} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \sum_{\alpha=1}^{n_a} \delta_i(s_\alpha) \chi_{A_k}(x_\alpha) \Phi_N(X, S) dX \\
&\stackrel{(*)}{=} \mu_0 \sum_{i,j=1}^{n_s} \sum_{k=1}^m \gamma_{ij}^{(k)} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) N_i^{(k)} \Phi_N(X, S) dX \\
&= \sum_{i,j=1}^{n_s} \sum_{k=1}^m \gamma_{ij}^{(k)} N_i^{(k)} \langle \Phi_M, \Phi_N \rangle \\
&= \begin{cases} \sum_{k=1}^m \sum_{i,j=1}^{n_s} \gamma_{ij}^{(k)} N_i^{(k)} \langle \Phi_N, \Phi_N \rangle, & \text{if } M = N, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

In (\*) we used that it holds  $\sum_{\alpha=1}^{n_a} \delta_i(s_\alpha) \chi_{A_k}(x_\alpha) = N_i^{(k)}$  for all  $(X, S)$  with  $\Phi_N(X, S) \neq 0$ . Analogously, we calculate the non-diagonal entries setting  $\gamma_{ii}(x) = 0$  such that we can

sum over all  $i, j$ :

$\mathcal{G}_{2,M,N}$

$$\begin{aligned}
&= \mu_0 \sum_{i,j=1}^{n_s} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \sum_{\alpha=1}^{n_a} \delta_j(s_\alpha) \gamma_{ij}(x_\alpha) \Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) dX \\
&= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i,j=1}^{n_s} \sum_{k=1}^m \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \delta_j(s_\alpha) \chi_{A_k}(x_\alpha) \gamma_{ij}(x_\alpha) \Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) dX \\
&= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i,j=1}^{n_s} \sum_{k=1}^m \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a-1}} \Phi_{M-E_j^{(k)}}(X^{-\alpha}, S^{-\alpha}) \delta_j(s_\alpha) \int_{\mathbb{X}} \chi_{A_k}(x_\alpha) \gamma_{ij}(x_\alpha) dx_\alpha \Phi_{N-E_i^{(k)}}(X^{-\alpha}, S^{-\alpha}) dX^{-\alpha} \\
&= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{i,j=1}^{n_s} \sum_{k=1}^m \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a-1}} \Phi_{M-E_j^{(k)}}(X^{-\alpha}, S^{-\alpha}) \delta_j(s_\alpha) \gamma_{ij}^{(k)} \int_{\mathbb{X}} \chi_{A_k}(x_\alpha) dx_\alpha \Phi_{N-E_i^{(k)}}(X^{-\alpha}, S^{-\alpha}) dX^{-\alpha} \\
&= \mu_0 \sum_{i,j=1}^{n_s} \sum_{k=1}^m \gamma_{ij}^{(k)} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a-1}} \Phi_M(X, S) \sum_{\alpha=1}^{n_a} \delta_j(s_\alpha) \chi_{A_k}(x_\alpha) \Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) dX \\
&\stackrel{(**)}{=} \mu_0 \sum_{i,j=1}^{n_s} \sum_{k=1}^m \gamma_{ij}^{(k)} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a-1}} \Phi_M(X, S) (N_j^{(k)} + 1) \Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) dX \\
&= \sum_{i,j=1}^{n_s} \sum_{k=1}^m \gamma_{ij}^{(k)} (N_j^{(k)} + 1) \langle \Phi_M, \Phi_{N+E_j^{(k)}-E_i^{(k)}} \rangle \\
&= \begin{cases} \gamma_{ij}^{(k)} (N_j^{(k)} + 1) \langle \Phi_M, \Phi_M \rangle, & \text{if } M = N + E_j^{(k)} - E_i^{(k)}, i \neq j, \\ 0, & \text{otherwise,} \end{cases} \\
&\stackrel{(2.9)}{=} \begin{cases} \gamma_{ij}^{(k)} N_i^{(k)} \langle \Phi_N, \Phi_N \rangle, & \text{if } M = N + E_j^{(k)} - E_i^{(k)}, i \neq j, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

In (\*\*) we used that it holds  $\sum_{\alpha=1}^{n_a} \delta_j(s_\alpha) \chi_{A_k}(x_\alpha) = N_j^{(k)} + 1$  for all  $(X, S)$  with  $\Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) \neq 0$ .

Combining the diagonal and non-diagonal part and using Corollary 3, we obtain

$$\hat{\mathcal{G}}_{NM} = \begin{cases} \gamma_{ij}^{(k)} N_i^{(k)}, & \text{if } M = N + E_j^{(k)} - E_i^{(k)}, i \neq j, \\ -\sum_{i,j=1}^{n_s} \sum_{k=1}^m \gamma_{ij}^{(k)} N_i^{(k)}, & \text{if } M = N, \\ 0, & \text{otherwise} \end{cases}$$

for the entries of the matrix  $\hat{\mathcal{G}}$ . □

## Second-order adoptions

For the second-order status-changes, we consider the adoption rate function  $f_{ij}^{(\alpha)}$  defined in (1.7). We assume that the interaction radius  $r$  is small compared to the diameters of the sets  $A_k$  such that second-order adoptions will mainly be caused by agents of the same subpopulation as these agents are located relatively close to each other in space. Near the boundaries of the spatial subsets also agents of different subpopulations can still be close enough to interact with each other, even in the case of a small interaction radius

$r > 0$ . In a metastable system, agents are sparsely positioned at the boundaries, and thus the probability of second-order adoptions due to cross-over interactions between different subpopulations approaches zero. More precisely, let

$$b_{kl} := \frac{\int_{\mathbb{X}^2} d_r(x_1, x_2) \chi_{A_k}(x_1) \chi_{A_l}(x_2) dx_1 dx_2}{\int_{\mathbb{X}^2} \chi_{A_k}(x_1) \chi_{A_l}(x_2) dx_1 dx_2} \quad (2.17)$$

for  $d_r$  given in (1.9) denote the conditional probability for two agents to have a physical distance closer than the interaction radius, given that they are located in the sets  $A_k$  and  $A_l$ , respectively. Given the coarse-grained system state  $N$  of the SMM, we define

$$\epsilon_{ij}^{(k)}(N) := c_{ij} \sum_{\substack{l=1 \\ l \neq k}}^m b_{kl} N_i^{(k)} N_j^{(l)} \quad (2.18)$$

as the equilibrium intensity rate for an adoption event  $i \rightarrow j$  to take place in subpopulation  $k$  because of a cross-over interaction with agents from a different subpopulation  $l \neq k$ , and

$$\hat{\gamma}_{ij}^{(k)} := c_{ij} b_{kk} \quad (2.19)$$

as the mesoscopic rate constant for adoptions within subpopulation  $k$ . Using these definitions, we obtain the following result.

**Theorem 39.** *For second-order adoptions with an ABM rate function  $f_{ij}^{(\alpha)}$  given by (1.7) and (1.8), the projected generator  $Q\hat{G}Q$  has the matrix representation  $\hat{\mathcal{G}}$  with*

$$\hat{\mathcal{G}}_{NM} = \begin{cases} \hat{f}_{ij}^{(k)}(N) + \epsilon_{ij}^{(k)}(N), & \text{if } M = N + E_j^{(k)} - E_i^{(k)}, i \neq j, \\ -\sum_{i,j=1}^{n_s} \sum_{k=1}^m \left( \hat{f}_{ij}^{(k)}(N) + \epsilon_{ij}^{(k)}(N) \right), & \text{if } M = N, \\ 0, & \text{otherwise} \end{cases}$$

where

$$\hat{f}_{ij}^{(k)}(N) := \hat{\gamma}_{ij}^{(k)} N_i^{(k)} N_j^{(k)}. \quad (2.20)$$

*Proof.* At first, we observe that for each  $i, j \in \mathbb{S}$ ,  $S \in \mathbb{S}^{n_a}$ ,  $\alpha \in \{1, \dots, n_a\}$ , and  $M \in \mathbb{M}_{n_a}$  it holds

$$\begin{aligned} & \delta_i(s_\alpha) \delta_j(s_\beta) \int_{\mathbb{X}^{n_a}} \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) d_r(x_\alpha, x_\beta) \Phi_M(X, S) dX \\ &= b_{kl} \delta_i(s_\alpha) \delta_j(s_\beta) \int_{\mathbb{X}^{n_a}} \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) \Phi_M(X, S) dX. \end{aligned} \quad (2.21)$$

This can be seen from the following calculation:

$$\begin{aligned} & \delta_i(s_\alpha) \delta_j(s_\beta) \int_{\mathbb{X}^{n_a}} \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) d_r(x_\alpha, x_\beta) \Phi_M(X, S) dX \\ &= \delta_i(s_\alpha) \delta_j(s_\beta) \int_{\mathbb{X}^{n_a-2}} \int_{\mathbb{X}^2} \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) d_r(x_\alpha, x_\beta) dx_1 dx_2 \Phi_{M-E_i^{(k)}-E_j^{(l)}}(X, S) dX \\ &\stackrel{(2.17)}{=} \delta_i(s_\alpha) \delta_j(s_\beta) \int_{\mathbb{X}^{n_a-2}} b_{kl} \int_{\mathbb{X}^2} \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) dx_1 dx_2 \Phi_{M-E_i^{(k)}-E_j^{(l)}}(X, S) dX \\ &= b_{kl} \delta_i(s_\alpha) \delta_j(s_\beta) \int_{\mathbb{X}^{n_a}} \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) dx_1 dx_2 \Phi_M(X, S) dX. \end{aligned}$$

We use the same decomposition (2.16) as in the proof for the first-order adoption rate functions. Let the rate function  $\gamma_{ij}$  be given by  $\gamma_{ij}(x_1, x_2) = c_{ij} \cdot d_r(x_1, x_2)$  for  $c_{ij} \geq 0$ , see Equation (1.8) for a definition, where  $c_{ii} = 0$ . Note that by setting  $c_{ii} = 0$  we can take the sum over all  $\alpha, \beta$  without the condition  $\beta \neq \alpha$  because it holds  $\delta_i(s_\alpha) \delta_j(s_\beta) c_{ij} = 0$  for all  $i, j$  in case of  $\alpha = \beta$ . We compute

$$\begin{aligned} \mathcal{G}_{1,M,N} &= \mu_0 \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \sum_{i,j=1}^{n_s} \sum_{\alpha=1}^{n_a} f_{ij}^{(\alpha)}(X, S) \Phi_N(X, S) dX \\ &= \mu_0 \sum_{\alpha,\beta=1}^{n_a} \sum_{i,j=1}^{n_s} c_{ij} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \delta_i(s_\alpha) \delta_j(s_\beta) d_r(x_\alpha, x_\beta) \Phi_N(X, S) dX. \end{aligned}$$

As we have  $\Phi_M \cdot \Phi_N = 0$  for  $M \neq N$ , we can follow  $\mathcal{G}_{1,M,N} = 0$  for  $M \neq N$ . For  $M = N$ , on the other hand, we have  $\Phi_M \cdot \Phi_N = \Phi_N$ , such that we can omit  $\Phi_M$  and get

$$\begin{aligned} \mathcal{G}_{1,M,N} &= \mu_0 \sum_{\alpha,\beta=1}^{n_a} \sum_{i,j=1}^{n_s} c_{ij} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \delta_i(s_\alpha) \delta_j(s_\beta) d_r(x_\alpha, x_\beta) \Phi_N(X, S) dX \\ &= \mu_0 \sum_{\alpha,\beta=1}^{n_a} \sum_{i,j=1}^{n_s} \sum_{k,l=1}^m c_{ij} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \delta_i(s_\alpha) \delta_j(s_\beta) \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) d_r(x_\alpha, x_\beta) \Phi_N(X, S) dX \\ &\stackrel{(2.21)}{=} \mu_0 \sum_{\alpha,\beta=1}^{n_a} \sum_{i,j=1}^{n_s} \sum_{k,l=1}^m c_{ij} b_{kl} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \delta_i(s_\alpha) \delta_j(s_\beta) \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) \Phi_N(X, S) dX \\ &= \mu_0 \sum_{i,j=1}^{n_s} \sum_{k,l=1}^m c_{ij} b_{kl} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \sum_{\alpha,\beta=1}^{n_a} \delta_i(s_\alpha) \delta_j(s_\beta) \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) \Phi_N(X, S) dX \\ &\stackrel{(*)}{=} \mu_0 \sum_{i,j=1}^{n_s} \sum_{k,l=1}^m c_{ij} b_{kl} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} N_i^{(k)} N_j^{(l)} \Phi_N(X, S) dX \\ &= \sum_{i,j=1}^{n_s} \sum_{k,l=1}^m c_{ij} b_{kl} N_i^{(k)} N_j^{(l)} \langle \Phi_N, \Phi_N \rangle, \end{aligned}$$

where (\*) is true because of  $\sum_{\alpha=1}^{n_a} \delta_i(s_\alpha) \chi_{A_k}(x_\alpha) = N_i^{(k)}$  and  $\sum_{\beta=1}^{n_a} \delta_j(s_\beta) \chi_{A_l}(x_\beta) = N_j^{(l)}$  for all  $(X, S)$  with  $\Phi_N(X, S) \neq 0$ . Using

$$f_{ij}^{(\alpha)}(X, S + ie_\alpha - je_\alpha) = c_{ij} \delta_j(s_\alpha) \sum_{\beta \neq \alpha} d_r(x_\alpha, x_\beta) \delta_j(s_\beta)$$

with  $c_{ij} = 0$  for  $i = j$ , we analogously get for the non-diagonal entries:

$$\begin{aligned}
\mathcal{G}_{2,M,N} &= \mu_0 \sum_{i,j=1}^{n_s} c_{ij} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \sum_{\alpha=1}^{n_a} \delta_j(s_\alpha) \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{n_a} \delta_j(s_\beta) d_r(x_\alpha, x_\beta) \Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) dX \\
&= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{n_a} \sum_{i,j=1}^{n_s} \sum_{k,l=1}^m c_{ij} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \Phi_M(X, S) \delta_j(s_\alpha) \delta_j(s_\beta) \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) d_r(x_\alpha, x_\beta) \Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) dX.
\end{aligned}$$

In case of  $M \neq N + E_j^{(k)} - E_i^{(k)}$  for all  $i, j = 1, \dots, n_s$  and all  $k = 1, \dots, m$  we get  $\mathcal{G}_{2,M,N} = 0$  because it holds  $\Phi_M \cdot \Phi_{N+E_j^{(k)}-E_i^{(k)}} = 0$  for all  $i, j, k$ . For  $M = N$ , we have  $\mathcal{G}_{2,M,N} = 0$  because of  $c_{ii} = 0$ . If, on the other hand,  $M = N + E_j^{(k)} - E_i^{(k)}$  for some  $i \neq j$  and  $k \in \{1, \dots, m\}$ , we have  $\Phi_M \cdot \Phi_{N+E_j^{(k)}-E_i^{(k)}} = \Phi_{N+E_j^{(k)}-E_i^{(k)}}$  and

$$\begin{aligned}
\mathcal{G}_{2,M,N} &= \mu_0 \sum_{\alpha=1}^{n_a} \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{n_a} \sum_{l=1}^m c_{ij} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \delta_j(s_\alpha) \delta_j(s_\beta) \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) d_r(x_\alpha, x_\beta) \Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) dX \\
&\stackrel{(2.21)}{=} \mu_0 \sum_{\alpha=1}^{n_a} \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{n_a} \sum_{l=1}^m c_{ij} b_{kl} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \delta_j(s_\alpha) \delta_j(s_\beta) \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) \Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) dX \\
&= \mu_0 \sum_{l=1}^m c_{ij} b_{kl} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} \sum_{\alpha=1}^{n_a} \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{n_a} \delta_j(s_\alpha) \delta_j(s_\beta) \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) \Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) dX \\
&\stackrel{(**)}{=} \mu_0 \sum_{l=1}^m c_{ij} b_{kl} \sum_{S \in \mathbb{S}^{n_a}} \int_{\mathbb{X}^{n_a}} (N_j^{(k)} + 1) N_j^{(l)} \Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) dX \\
&= \sum_{l=1}^m c_{ij} b_{kl} (N_j^{(k)} + 1) N_j^{(l)} \langle \Phi_{N+E_j^{(k)}-E_i^{(k)}}, \Phi_{N+E_j^{(k)}-E_i^{(k)}} \rangle \\
&\stackrel{(2.9)}{=} \sum_{l=1}^m c_{ij} N_i^{(k)} N_j^{(l)} b_{kl} \langle \Phi_N, \Phi_N \rangle.
\end{aligned}$$

In (\*\*) we used that

$$\sum_{\alpha=1}^{n_a} \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{n_a} \delta_j(s_\alpha) \delta_j(s_\beta) \chi_{A_k}(x_\alpha) \chi_{A_l}(x_\beta) = (N_j^{(k)} + 1) N_j^{(l)}$$

holds for all  $(X, S)$  with  $\Phi_{N+E_j^{(k)}-E_i^{(k)}}(X, S) \neq 0$ . For  $k \neq l$  this is clear because for these  $(X, S)$  there must be  $N_j^{(k)} + 1$  agents of status  $j$  located in subset  $A_k$  and  $N_j^{(l)}$  agents of status  $j$  in subset  $A_l \neq A_k$ . For  $l = k$  we obtain  $(N_j^{(k)} + 1) N_j^{(k)}$ , which is the number of pairs of different agents both of status  $j$  and being located in  $A_k$ . Dividing by  $\langle \Phi_N, \mathbb{1} \rangle = \langle \Phi_N, \Phi_N \rangle$  (see again Corollary 3) and combining diagonal and non-diagonal

entries, we find that the matrix  $\hat{\mathcal{G}}$  has the entries

$$\hat{\mathcal{G}}_{NM} = \begin{cases} \sum_{l=1}^m c_{ij} b_{kl} N_i^{(k)} N_j^{(l)}, & \text{if } M = N + E_j^{(k)} - E_i^{(k)}, i \neq j, \\ -\sum_{i,j=1}^{n_s} \sum_{k,l=1}^m c_{ij} b_{kl} N_i^{(k)} N_j^{(l)}, & \text{if } M = N, \\ 0, & \text{otherwise} \end{cases}$$

By using Equations (2.18), (2.19) and (2.20) we complete the proof.  $\square$

### Projection error

Theorem 39 shows how the projected rate functions of second-order adoptions can be decomposed into: (1) a part coming from interactions between agents of the same subpopulation with adoption rate functions given by  $\hat{f}_{ij}^{(k)}$ ; and (2) a part coming from adoptions that take place between agents of different subpopulations. As discussed above, in a metastable system and for a metastable full partition  $A_1, \dots, A_m$ , the probability of cross-over interactions is small and thus the value of  $\epsilon_{ij}^{(k)}(N)$  will be negligibly small. As we will see in the following subsection for the application to the guiding Example 2.2.5, the error introduced by omitting cross-over interactions is often orders of magnitude smaller than the discretization error of the Galerkin projection.

### Spatio-temporal master equation

In the SMM, therefore, we consider only the first type (1) of second-order adoptions, i.e. we assume that there are no (cross-over) adoptions between agents of different subpopulations. Using the results of Theorems 37-39, the equation for the evolution of the SMM given by (2.3) and (2.4) can be written as the following spatio-temporal master equation:

$$\begin{aligned} \frac{dP(N, t)}{dt} = & - \sum_{\substack{k,l=1 \\ k \neq l}}^m \sum_{i=1}^{n_s} \lambda_i^{(kl)} N_i^{(k)} P(N, t) \\ & + \sum_{\substack{k,l=1 \\ k \neq l}}^m \sum_{i=1}^{n_s} \lambda_i^{(kl)} (N_i^{(k)} + 1) P(N + E_i^{(k)} - E_i^{(l)}, t) \\ & - \sum_{i,j=1}^{n_s} \sum_{k=1}^m \hat{f}_{ij}^{(k)}(N) P(N, t) \\ & + \sum_{i,j=1}^{n_s} \sum_{k=1}^m \hat{f}_{ij}^{(k)}(N + E_i^{(k)} - E_j^{(k)}) P(N + E_i^{(k)} - E_j^{(k)}, t), \end{aligned} \tag{2.22}$$

where the first two terms on the right-hand side refer to the change caused by the spatial transitions between subpopulations defined by the rate matrix  $\mathcal{L}$ , while the other two

terms are describing the change through status adoptions within the subpopulations described by the operator  $\mathcal{G}$ .<sup>2</sup>

### Relation to models for chemical reaction kinetics

Using the definitions of the functions  $\hat{f}_{ij}^{(k)}$  given in Theorems 38 and 39, the interaction propensities are consistent with the standard law of mass-action from the chemical context [79]. This is due to the fact that we assume that the agents interact independently of each other (and of the overall system state) – which we do by choosing the ABM adoption rate functions according to Equations (1.6) and (1.7). As for the spatial dynamics, state-of-the art metapopulation models [76] typically assume that the commuting flow between two subpopulations  $k$  and  $l$  is of the form  $\lambda_i^{(kl)}(N_k)^a(N_l)^b$  for exponents  $a, b \geq 0$  which tune the dependence with respect to each subpopulation size. In our setting, we assume that the spatial movement of each agent is independent of the population sizes, which corresponds to setting  $a = 1$  and  $b = 0$ .

### Extension to core set approach

Instead of a partition of  $\mathbb{X}$  into  $m$  metastable sets, we can also consider a decomposition based on  $m$  core sets  $C_1, \dots, C_m$ . The core sets should be chosen such that all core regions of the corresponding sets of the metastable partition are included and the transition regions with a low invariant measure are not covered by the sets [69, 74, 75]. The subpopulations of the coarse grained system state can then be interpreted either in the sense of a milestoning process as was described in Section 2.1.3 or in the probabilistic sense of the associated Galerkin projection based on the isocommittor functions defined in (2.1). In the milestoning interpretation, agents within subpopulation  $k$  are those who visited the core set  $C_k$  last. The associated Galerkin projection however associates the agents of the ABM to a subpopulation of the projected process according to the committor probabilities. Based on the position  $x$  the committor function takes the value  $q_k(x)$ , which describes the probability of an individual agent to have visited the core set  $C_k$  last.

For the construction of a core set SMM, where the system state can be interpreted in the milestoning sense we need to adapt our results for the matrix representations [16]. The spatial transition rates  $\lambda_i^{(kl)}$  correspond in this case to the transition rates of a milestoning process between the core sets instead of the transition rates between the partition sets and are sampled from the trajectory data as described in Section 2.1.3. The Equation (2.15) for first-order adoptions is adjusted to

$$\gamma_{ij}^{(k)} := \frac{\langle \gamma_{ij}, q_k \rangle_{\mathbb{X}}}{\langle q_k, \mathbb{1} \rangle_{\mathbb{X}}} = \frac{\int_{\mathbb{X}} \gamma_{ij}(x) q_k(x) dx}{\int_{\mathbb{X}} q_k(x) dx} \quad (2.23)$$

---

<sup>2</sup>In the second line of (2.22), we need the rate to go from  $M := N + E_i^{(k)} - E_i^{(l)}$  to the given  $N$ . By Theorem 37 we know that this rate is given by  $\lambda_i^{(kl)} M_i^{(k)} = \lambda_i^{(kl)} (N_i^{(k)} + 1)$ .

and the interaction probabilities for the second-order adoptions are modified to

$$b_{kl} := \frac{\int_{\mathbb{X}^2} d_r(x_1, x_2) q_{C_k}(x_1) q_{C_l}(x_2) dx_1 dx_2}{\int_{\mathbb{X}^2} q_{C_k}(x_1) q_{C_l}(x_2) dx_1 dx_2}. \quad (2.24)$$

Compared to the full partition results we replaced the characteristic functions of the partition sets with the committor functions of the respective core sets. Since the value of the scalar product between two different committor functions is small we still have a negligibly small error  $\epsilon_{ij}^{(k)}(N)$  due to adoptions between different subpopulations, but in addition also a reduced error for the spatial discretization.

**Remark 2** (Approximation quality). The step from the full-scale ABM to the SMM (2.22) involves two approximations: the discretization error originating from the Galerkin projection and the error resulting from neglecting the cross-over interactions between different spatial domains. While the latter error can be easily monitored by estimating the neglected cross-over rates, the discretization error is more difficult to control. For cases where the spatial motion exhibits metastability, estimating the discretization error for the mobility process is possible (see [75] for the core set approach), but requires sufficient ABM simulation data.

So far, we have been referring to the metastability of the mobility process when we talked about metastability in the ABM. For the SMM to be a good approximation of the ABM, there must also be a time scale separation between the mobility and adoption dynamics, in the sense that mixing within a subpopulation occurs on a faster time scale than adoption events. If this is not the case, then the assumption that agents are well mixed within a subpopulation is no longer valid with respect to the adoption process, since there may be dependencies for the adoption rates on the exact position within the subpopulation that are no longer reflected by the averaged interaction probabilities  $b_{kl}$ .

## 2.2.5 Application to the Guiding example

**Example 6 (continued).** Given the double-well potential shown in Figure 1.10 and using the Markov state model approach, we partition the space  $\mathbb{X} = \mathbb{R}^2$  of movement into the two core sets  $C_1 = (-\infty, -0.5) \times \mathbb{R}$  and  $C_2 = (0.5, \infty) \times \mathbb{R}$  and the transition region  $\mathbb{X} \setminus (C_1 \cup C_2)$ . The system state of the SMM is a  $2 \times 2$  matrix and the initial system state is chosen with values  $N_1^{(1)} = 49, N_1^{(2)} = 1, N_2^{(1)} = 50$  and  $N_2^{(2)} = 0$ , which is also the projected initial state of the ABM. The jump rates  $\lambda_i^{(12)}$  and  $\lambda_i^{(21)}$  between the two subpopulations are the transition rates between  $C_1$  and  $C_2$  and are estimated using the maximum likelihood estimator of Section 2.1.3 for the milestone transitions of the ABM trajectory data. The interaction probabilities  $b_{11}$  and  $b_{22}$  for the scaling of the projected adoption rate functions for the SMM  $\hat{f}_{12}^{(1)}(N) = c_{12}b_{11}N_1^{(1)}N_2^{(1)}$  and  $\hat{f}_{12}^{(2)}(N) = c_{12}b_{22}N_1^{(2)}N_2^{(2)}$  are estimated from trajectory data as well.

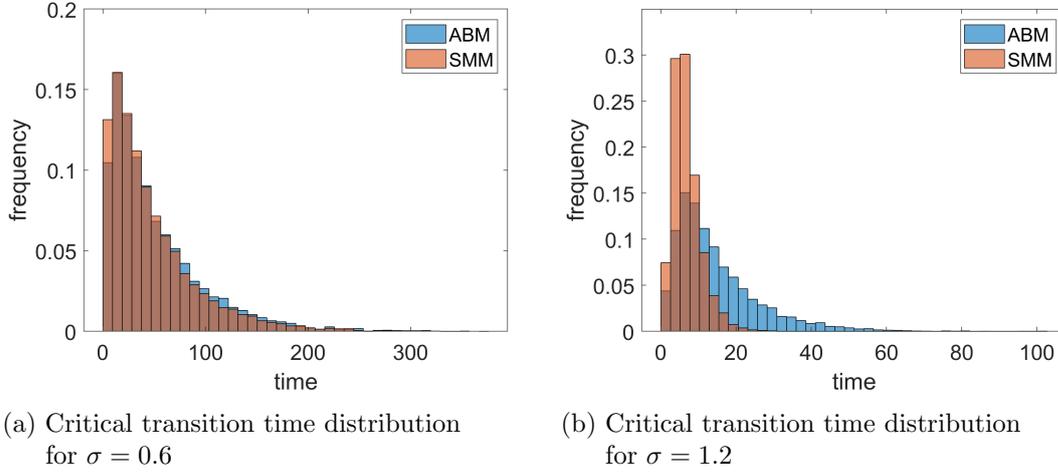


Figure 2.2: Distribution of the critical transition time for the ABM (blue) and the SMM (orange) given in Example 6, sampled over 10000 MC-simulations. The overlap of the two distributions is colored brown. For small  $\sigma$  the distribution is very well matched, while for larger  $\sigma$  the critical transition happens faster in the SMM.

We analyze the quality of the SMM approximation for two different values of the diffusion constant,  $\sigma = 0.6$  and  $\sigma = 1.2$ , where the first case is more metastable than the other. We compare the SMM process with the projected ABM dynamics with respect to the temporal distribution of the *critical transition event* given by the first agent with status 2 switching from one subpopulation to the other (for the projected ABM, this means that an agent with status 2 who last visited core set  $C_1$  reaches core set  $C_2$  for the first time), see Figure 2.2. Since this transition has a large impact on the overall dynamics, we consider the overall approximation error to be small if the difference in the distribution of

this critical transition event time is small. We observe that for smaller  $\sigma$  the approximation is better due to an increase in the metastability of the dynamics and consequently a better approximation of the exit time of a subpopulation by an exponential holding time.

This becomes even clearer when comparing the temporal evolution of the average number of agents with status 2 in the two subpopulations, see Figure 2.3. For the smaller value  $\sigma = 0.6$ , the first moments agree very well (Figure 2.3a). In contrast, for the larger diffusion constant  $\sigma = 1.2$  there is a significant difference between the model results regarding these first moments (Figure 2.3b). This is due to the fact that the approximation quality of the Markov state model is worse because the diffusion process is less metastable and thus the first spatial transitions in the SMM happen too fast on average. The deviation of the spatial transition dynamics is also the main contributor to the approximation error of the critical transition event time, as the expected evolution of adoptions within the initial subpopulation is still well approximated even in the less metastable case. This is also because the mixing within the subpopulations is considerably faster than the adoption process in both cases. Numerical experiments with a strongly increased value for the constant  $c_{12}$  resulted in the SMM overestimating the frequency of adoption events, since in the projected ABM adoptions happened almost instantaneously whenever two agents were within the interaction radius of each other, and thus the agents could no longer be considered well mixed with respect to the adoption dynamics. While in the SMM for this scenario all agents have the same probability of interacting with each other, this was not the case in the ABM due to the relatively slow mixing compared to the fast adoptions.

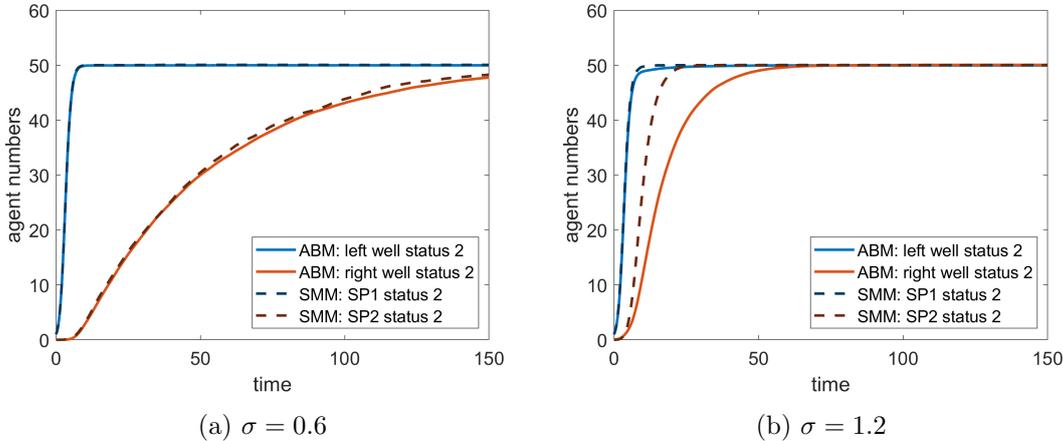


Figure 2.3: Comparison of the time-dependent mean population size of agents with status 2 over 10000 MC-simulations for the projected ABM and the corresponding stochastic metapopulation model. The projection is based on a core set approach with  $C_1$  and  $C_2$  given in Example 6 and defining two subpopulations denoted by SP1 and SP2. In scenario (a) the solid blue and dashed blue line are indistinguishable due to the good approximation quality.

## 2.3 Piecewise-Deterministic Metapopulation Model

For metapopulation models with a large population of interacting agents, the stochastic simulation of the stochastic metapopulation dynamics becomes computationally very expensive because it tracks every single adoption event. In this case, a further model reduction can be very useful to reduce the simulation effort. Since the number of agents in each subpopulation is large, we can apply standard convergence results and approximate the jump process describing the internal adoption dynamics, for which we assume a high intensity rate, by a deterministic evolution equation [66]. Such approximations, based on the law of large numbers, are well known in the context of chemical reaction systems, where they are used to reduce model complexity for systems with large molecular populations [79]. On the other hand, we assume that the spatial transition events have a low intensity rate and are comparatively rare events that occur only after long waiting times. To reduce the computational complexity of the simulations while preserving the discrete, stochastic nature of the transition events between the subpopulations, we approximate the overall dynamics by a piecewise deterministic Markov process, see [30, 80, 81] for details. This model will be called piecewise-deterministic metapopulation model (PDMM).

In this section, we will first introduce the ODE approximation to go from an SMM to a PDMM, and afterwards again apply the method to our guiding example and elaborate on the approximation quality.

### 2.3.1 ODE Approximation for the PDMM

The stochastic process  $(\mathbf{N}(t))_{t \in \mathbb{T}}$  given by (2.22) can be rewritten in a pathwise notation of the form

$$\begin{aligned} \mathbf{N}(t) = \mathbf{N}(0) &+ \sum_{\substack{k,l=1 \\ k \neq l}}^m \sum_{i=1}^{n_s} \mathcal{P}_i^{(kl)} \left( \int_0^t \lambda_i^{(kl)} \mathbf{N}_i^{(k)}(s) ds \right) (E_i^{(l)} - E_i^{(k)}) \\ &+ \sum_{i,j=1}^{n_s} \sum_{k=1}^m \mathcal{R}_{ij}^{(k)} \left( \int_0^t \hat{f}_{ij}^{(k)}(\mathbf{N}(s)) ds \right) (E_j^{(k)} - E_i^{(k)}), \end{aligned} \quad (2.25)$$

where  $\mathcal{P}_{ij}^{(k)}$  and  $\mathcal{R}_i^{(kl)}$  refer to independent, unit-rate Poisson processes [80, 82].

Assuming that the jumps induced by the Poisson processes  $\mathcal{P}_{ij}^{(k)}$ , which refer to the spatial transitions between metastable domains, occur much less frequently than the jumps induced by the Poisson processes  $\mathcal{R}_i^{(kl)}$ , which refer to the adoption dynamics within the subpopulations, we can apply standard convergence results for Markov processes [21] in order to approximate the stochastic dynamics given by the second line of Equation (2.25) by deterministic dynamics and obtain the PDMM process  $(\hat{\mathbf{N}}(t))_{t \in \mathbb{T}}$  given by the equation

$$\begin{aligned} \hat{\mathbf{N}}(t) = \hat{\mathbf{N}}(0) &+ \sum_{k,l=1, k \neq l}^m \sum_{i=1}^{n_s} \mathcal{P}_i^{(kl)} \left( \int_0^t \lambda_i^{(kl)} \hat{\mathbf{N}}_i^{(k)}(s) ds \right) (E_i^{(l)} - E_i^{(k)}) \\ &+ \sum_{i,j=1}^{n_s} \sum_{k=1}^m \int_0^t \hat{f}_{ij}^{(k)}(\hat{\mathbf{N}}(s)) ds (E_j^{(k)} - E_i^{(k)}). \end{aligned} \quad (2.26)$$

The first line is almost identical to the SMM description, as we keep the rare spatial transitions stochastic, while the second line now refers to an ODE system for modeling the local adoption dynamics of the population, similar to [81]. It is well-known [68, 79] that the relative error produced by this approximation (relative with respect to the population size) decreases as the number of agents increases. Therefore, we will consider a finite population, which is sufficiently large for the approximation to be reasonable. The ODE approximation of the PDMM can be interpreted as the expected evolution of the fast adoption process between two rare spatial transition events.

### 2.3.2 Application to the Guiding Example

**Example 6 (continued).** For our guiding example of two-status dynamics in a double-well potential, the deterministic status-adoption dynamics from the second line of (2.26) is given by

$$\hat{\mathbf{N}}(t_0 + \tau) = \hat{\mathbf{N}}(t_0) + \sum_{k=1}^m \int_{t_0}^{t_0 + \tau} \hat{f}_{12}^{(k)}(\hat{\mathbf{N}}(s))(E_2^{(k)} - E_1^{(k)}) ds$$

for  $\tau < t_1 - t_0$ , where  $t_0, t_1$  denote the time points of two subsequent stochastic transition events induced by the first line of (2.26). Using the definition (2.20) of  $\hat{f}_{12}^{(k)}$ , we get the following ODE for the number  $\hat{\mathbf{N}}_2^{(k)}$  of agents in subpopulation  $k$  having status 2:

$$\frac{d\hat{\mathbf{N}}_2^{(k)}(t)}{dt} = \hat{\gamma}_{12}^{(k)} \cdot \hat{\mathbf{N}}_1^{(k)}(t) \hat{\mathbf{N}}_2^{(k)}(t) \quad (2.27)$$

for  $t_0 < t < t_1$ . Let  $n_0^{(k)} := \hat{\mathbf{N}}_1^{(k)}(t_0) + \hat{\mathbf{N}}_2^{(k)}(t_0)$  denote the total number of agents in subpopulation  $k$  at time  $t_0$ . Between two transition events this number is constant, so we can substitute  $\hat{\mathbf{N}}_1^{(k)}(t) = n_0^{(k)} - \hat{\mathbf{N}}_2^{(k)}(t)$  in Equation (2.27) to arrive at

$$\frac{d\hat{\mathbf{N}}_2^{(k)}(t)}{dt} = \hat{\gamma}_{12}^{(k)} \hat{\mathbf{N}}_2^{(k)}(t) \left( n_0^{(k)} - \hat{\mathbf{N}}_2^{(k)}(t) \right).$$

This equation is a logistic ODE, which is a special case of Bernoulli ODEs for which an analytical solution is known. The solution is given by the logistic function, that is, we obtain an analytical solution

$$\hat{\mathbf{N}}_2^{(k)}(t) = n_0^{(k)} \left( 1 + e^{-\hat{\gamma}_{12}^{(k)} n_0^{(k)} (t_0) t} \left( n_0^{(k)} - \hat{\mathbf{N}}_2^{(k)}(t_0) \right) \right)^{-1}$$

for  $t_0 < t < t_1$ . Treating the diffusive transitions between the subpopulations as stochastic events which induce jumps in the state  $\hat{\mathbf{N}}$  of the PDMM process, we obtain trajectories as depicted in Figure 2.4 (b).

The main difference between the trajectories (see Figure 2.4) is that in the stochastic metapopulation model we have many discontinuous jumps while in the PDMM only a few remain. The error in estimating the critical event time induced by the PDMM approximation (see Figure 2.5) is small compared to the error arising from the spatial discretization (see Figure 2.2). Even though our population size of 100 individuals is not very large, the critical transition time distribution of the SMM is already well approximated by the PDMM for both choices of  $\sigma$ .

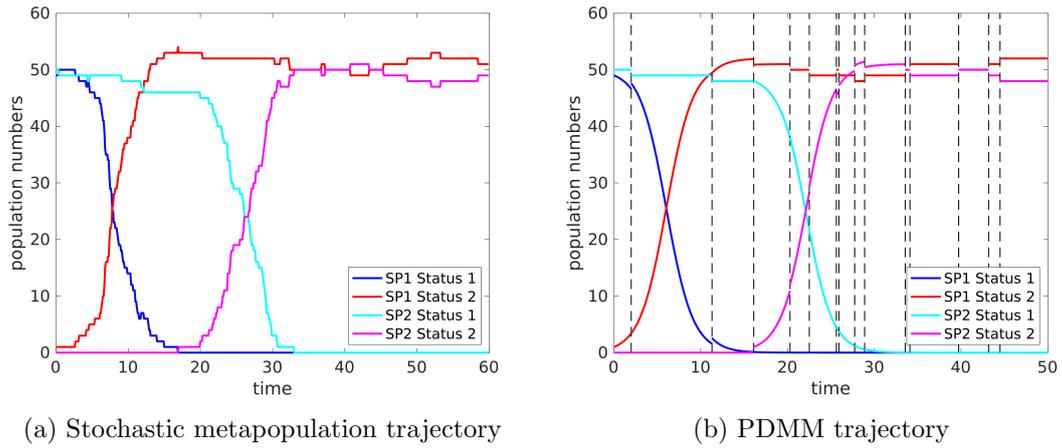


Figure 2.4: Comparison between (a) SMM and (b) PDMM trajectories for  $\sigma = 1.2$ . In (b) the rare jump events are marked by vertical dotted lines.

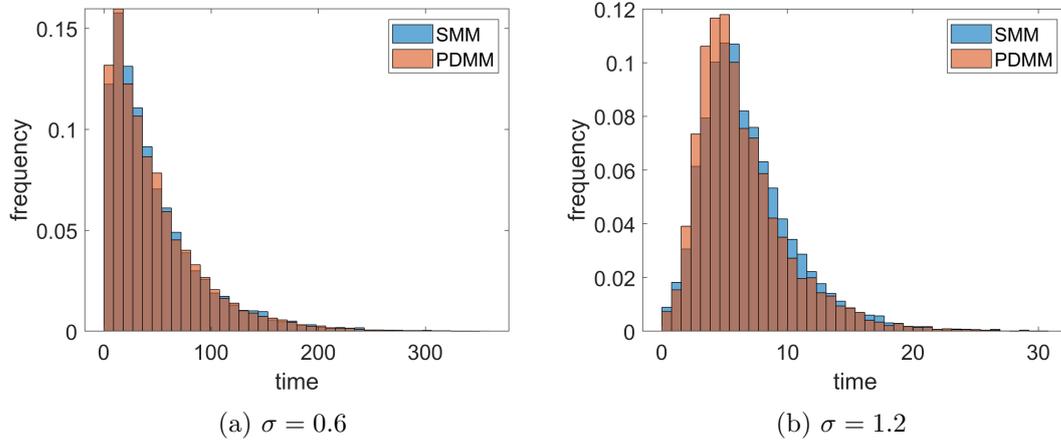


Figure 2.5: Distribution of the critical transition time for SMM (blue) and PDMM (orange) sampled over 10000 MC-simulations. The overlap of the two distributions is colored brown. For both values of  $\sigma$  the distribution is very well matched.

## 2.4 Simulation and Computational Efficiency

In this section, we look at the methods for simulating the two metapopulation model approaches presented in the previous section and also used for the realizations of the guiding example. We will compare the computational effort of the different simulation approaches, especially for the simulation runs of the guiding example.

### 2.4.1 Simulation of the SMM

For the stochastic metapopulation model, which is a pure jump process, we implement the sampling with the stochastic simulation algorithm, which produces statistically exact realizations of the process, without any numerical approximation error [55]. The pseudocode for applying the method to the SMM can be written as follows:

---

**Algorithm 7:** Stochastic simulation algorithm for the SMM

---

```

1 initialize time  $t = 0$ , system state  $\mathbf{N}(0)$  and time horizon  $T$ ;
2 while  $t < T$  do
3   # compute the total jump intensity for both transition and adoption events
4   set  $\Lambda := \sum_{k,l=1}^m \sum_{i=1}^{n_s} \lambda_i^{(kl)} \mathbf{N}_i^{(k)}(t) + \sum_{k=1}^m \sum_{i,j=1}^{n_s} \hat{f}_{ij}^{(k)}(\mathbf{N}(t))$ ;
5   draw waiting time  $\tau \sim \text{Exp}(\Lambda)$ ;
6   set Markov kernel  $\mathcal{Q}$  with
7   #probability for spatial transition events
8    $\mathcal{Q}(\mathbf{N}(t), \mathbf{N}(t) + E_i^{(l)} - E_i^{(k)}) := \frac{\lambda_i^{(kl)}}{\Lambda}$  and
9   #probability for adoption events
10   $\mathcal{Q}(\mathbf{N}(t), \mathbf{N}(t) + E_j^{(k)} - E_i^{(k)}) := \frac{\hat{f}_{ij}^{(k)}(\mathbf{N}(t))}{\Lambda}$ ;
11  draw a state change vector  $v$  according to  $\mathcal{Q}$ ;
12  # state update
13   $\mathbf{N}(t + \tau) = \mathbf{N}(t) + v$ ;
14  # time update
15   $t = t + \tau$ ;
16 end
Result:  $(\hat{\mathbf{X}}(t))_{t \leq T}$ 

```

---

## 2.4.2 Simulation of the PDMM

In order to simulate a PDMM process given by (2.26) one has to simultaneously integrate the deterministic flow of the ODE part and the state-dependent intensity rate functions  $\lambda_i^{(kl)}$  for the stochastic jumps, see [82]. Using the event-based simulation method for PDMPs from Section 1.3.3 one can determine the time point of the next stochastic jump while updating the system with the forward Euler scheme. Applied to the situation of the PDMM the simulation algorithm can be written as follows.

---

**Algorithm 8:** Event-based simulation for the PDMM

---

```

1 initialize time  $t = 0$  and state  $\hat{N}(0)$ ;
2 choose a time step  $\Delta t$  and time horizon  $T$ ;
3 draw  $\tau' \sim \text{Exp}(1)$  # exponentially distributed with rate 1;
4 while  $t < T$  do
5   # compute the jump intensity rate for spatial transitions
6   set  $\Lambda := \sum_{\substack{k,l=1 \\ k \neq l}}^m \sum_{i=1}^{n_s} \lambda_i^{(kl)} \hat{N}_i^{(k)}(t)$ ;
7   if  $\Lambda \Delta t > \tau'$  then
8     # jump event
9     set Markov kernel  $\mathcal{Q}$  with
10    #probability for spatial transition events
11     $\mathcal{Q}(\hat{N}(t), \hat{N}(t) + E_i^{(l)} - E_i^{(k)}) := \frac{\lambda_i^{(kl)}}{\Lambda}$ ;
12    draw a state change vector  $v$  according to  $\mathcal{Q}$ ;
13    # state update
14     $\hat{N}(t + \frac{\tau'}{\Lambda}) = \hat{N}(t) + \sum_{k=1}^m \sum_{i,j=1}^{n_s} (E_j^{(k)} - E_i^{(k)}) \hat{f}_{ij}^{(k)}(\hat{N}(t)) \frac{\tau'}{\Lambda} + v$ ;
15    # time update
16     $t = t + \frac{\tau'}{\Lambda}$ ;
17    draw new  $\tau' \sim \text{Exp}(1)$ ;
18  else
19    # state update
20     $\hat{N}(t + \Delta t) = \hat{N}(t) + \sum_{k=1}^m \sum_{i,j=1}^{n_s} (E_j^{(k)} - E_i^{(k)}) \hat{f}_{ij}^{(k)}(\hat{N}(t)) \Delta t$ ;
21    # time update
22     $t = t + \Delta t$ ;
23     $\tau' = \tau' - \Lambda \Delta t$ ;
24  end
25 end
Result:  $(\hat{X}(t))_{t \leq T}$ 

```

---

### 2.4.3 Effort Comparison

In addition to the details of the mobility process, which already make the ABM effort several orders of magnitude higher than the reduced models, we need to perform a range search in each time step to evaluate the adoption rate functions. The pairwise comparison of distances between agents scales quadratically with the number of agents  $n_a$ , but even for neighborhood computations using k-d-trees the computational effort still scales superlinearly with  $\mathcal{O}(n_a \log(n_a))$  [83, 84].

The simulation effort of the SMM is much lower than for ABM simulations, since the diffusive mobility process is reduced to a jump process with a low intensity rate, but still scales linearly with the number of agents, since each adoption event requires a separate simulation step. This means that for large populations, sampling the quantities of interest may still be infeasible due to high computational cost.

The numerical effort of the PDMM simulations does not scale significantly with the number of agents for our example, and thus, among the approaches considered in this work, the PDMM is the most efficient choice for simulating systems with high populations or large numbers of agents. For small numbers of agents, the PDMM effort is higher compared to the SMM because the critical transition event occurs significantly later and thus more time steps are needed in the PDMM computation. See Figure 2.6 for a comparison of the computational effort for the three modeling approaches.

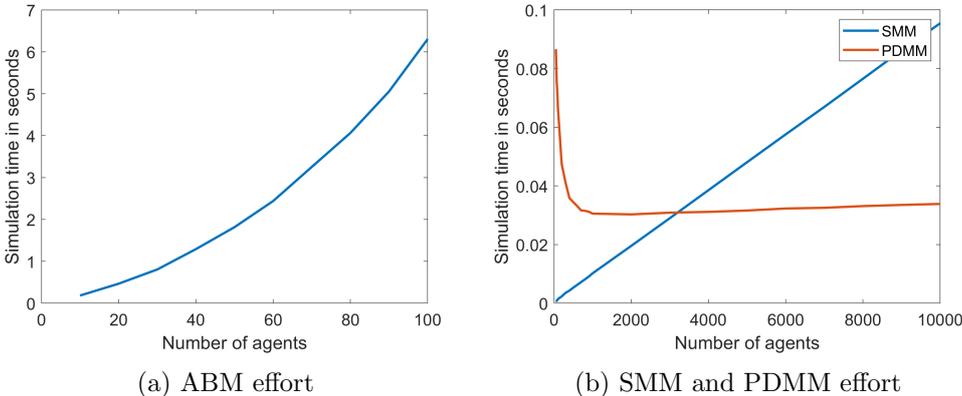


Figure 2.6: Numerical effort for the simulations of the guiding example system of Ex. 6 for different choices of agent numbers  $n_a$ , depending on the number of agents. Even for low agent numbers, the approximate models are at least two orders of magnitude more efficient.

Overall, the reduction in effort from the PDMM approximation of ABMs is immense and can allow sampling of large ABM sets that would otherwise be infeasible. This is especially the case when we consider models based on real-world data, as we do in the next chapter.

## 3 Modeling of Spreading Processes

In this chapter, we will focus on applying the concepts from Chapter 1 to formalize and construct agent-based models for real-world processes. We will also look at cases where the concepts from Chapter 2 can be applied to either derive a reduced model that preserves important properties of the ABM or to directly construct a metapopulation model for the process. All applications are based on real-world data, and we will illustrate how the data can be incorporated into the construction and calibration of the models. Unfortunately, due to the complexity of the models, formal analysis will not be feasible, and we will instead focus on methods for computational analysis. While in the previous chapter we discussed how metastability is a prerequisite for deriving a reduced model with good approximation quality, in this chapter we will focus more on how to identify metastable sets and discuss approaches that use clustering methods.

The processes that we are interested in are spreading processes in general, such as the dispersal of a species in physical space or spreading processes within a human population, such as the transmission of a disease or information. While different types of spreading processes each have their distinct dynamical properties, there are two general concepts that are key features for many of them. The first is mobility, and the second is interaction. For the dispersal of a species it is clear that in order to spread to a new location in physical space the members of that species must be mobile themselves or be able to produce mobile vectors for reproduction, such as pollen or spores [81]. The ability of a species to persist in a location depends on interactions with other members of the same species, the environment or other competing species [85]. For information to propagate within human populations, interactions in the form of some kind of communication are mandatory, and usually physical proximity is either required or a major influence. Therefore, mobility plays an important role in these processes. Even more so in the case of infectious disease spreading, where actual physical contact is required to transmit a virus or other vector from one individual to another. Thus, in modeling spreading processes, one must capture the influence and type of mobility involved in the original process as well as the interaction patterns when designing the model dynamics.

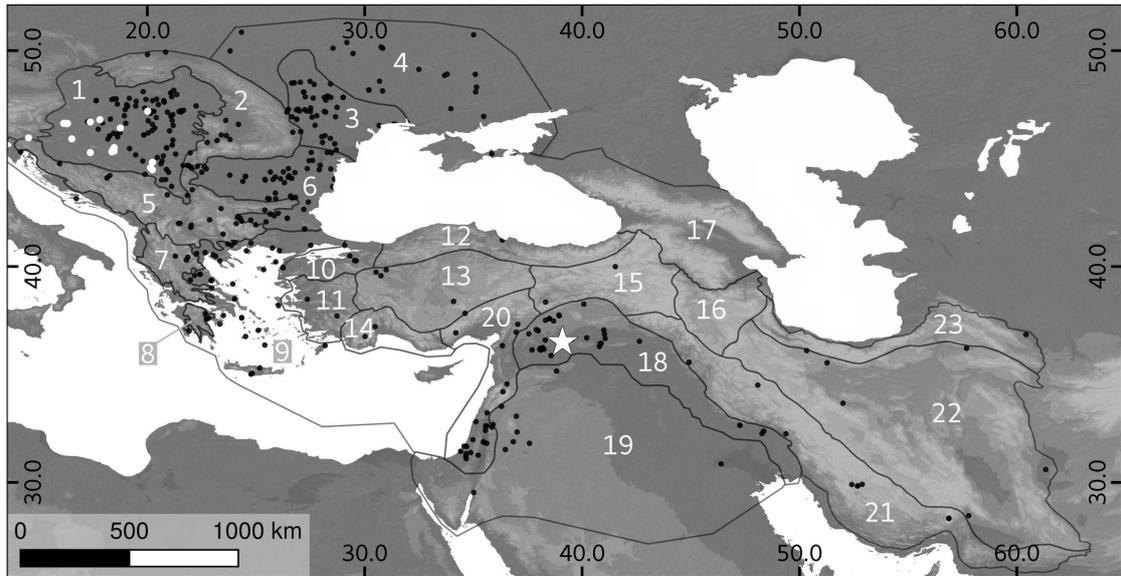
In total, we will discuss three models for spreading processes in more detail. In the first section, we will consider an agent-based model for the spreading of the woolly sheep in ancient times, which is an example for innovation spreading within and between human populations. The model that we discuss in the second section is an ABM for the mobility and cultural evolution of hunter-gatherer societies in prehistoric times. These are two examples of applications in the prehistoric context in which the collection and integration of data as well as the validation of the model are challenging tasks. We

discuss where the difficulties lie for our examples and how the model results can be interpreted. Afterwards, we consider in the third section the spreading of infectious diseases in present day societies. While we also briefly discuss agent-based approaches at the end of the section, our focus will be the metapopulation level. For each of the presented models, we will discuss the assumptions about the real-world processes that we make and which level of detail the model should have depending on the research question. We will then derive a formal description of each model in terms of Markov processes and apply the simulation algorithms of Chapter 1 to provide an implementation. The simulation results of each model are discussed in terms of a computational analysis with focus on the sampling of rare events such as critical transitions, macroscopic variables that quantify the complex system states and the identification of mesoscopic patterns such as clustering of agents. We also discuss the applicability of the model reduction framework from Chapter 2 for each of the applications and how it could be adapted for more general dynamics.

### 3.1 The Spreading of the Woolly Sheep in Ancient Times

The first application of the previously discussed approaches to a real-world scenario will be the spreading of the wool-bearing sheep in ancient times, which has also been discussed by me and co-authors in [14, 15]. The topic is of interest to archaeologists, as the woolly sheep represents an important innovation for the manufacturing of textiles at the time and subsequently also had a huge impact on the socio-economic development of past societies [86, 87]. The modeling prehistoric processes faces several additional difficulties compared to modern day settings. One challenge is the limited availability of data, which tends to be sparse, indirect and uncertain [88]. The localization, dating and categorization of archaeological findings already involves estimation techniques and the possibility of measurement errors that must be taken into account. Another challenge is that it is not possible to generate additional data through experiments or to replicate the prehistoric process. While it is possible to perform physical experiments, for example, to estimate the efficiency of ancient mining or smelting techniques [89], or to evaluate the suitability of certain materials for tool making [90], we cannot recreate an ancient society in a similar way. Thus, many of the state-of-the-art approaches based on data fitting and validation [91] cannot be directly applied, and reconstruction of the original process is generally not feasible. Instead, we want to conduct a rational modeling approach that utilizes expert knowledge and additional assumptions to compensate for the gaps in the data. The resulting model can then be used as a tool to test hypotheses and illustrate the characteristics of the prehistoric process.

The area of interest for our study is the geographical region that includes the Near East and Southeastern Europe (see Figure 3.1). This region includes the presumed origin of the woolly sheep in present-day Northern Syria where the earliest archaeological evidence of wool production is located [92]. The available archaeological data for our study region includes a set of spindle whorl findings in Southeastern Europe and a set of ovicaprid bone finds spread throughout the region but with the majority of the data points also being located in Southeastern Europe. Both data sets are visualized in Figure 3.1. While the spindle whorls can be linked to wool processing within the area of the site, we can not make the same connection for the bone findings. From the analysis of the bones we can only infer that sheep or goats have been present in the area but not whether the sheep that left behind the bone was wool-bearing or not. Thus, the only facts that we can infer with relative certainty from the data is the starting location and time of the spreading process as well as the presence of the innovation in the northwestern edge of our study area approximately 2000 years later. While the presence of ovicaprid bones could be interpreted as evidence that the environment has been suitable for keeping sheep the amount of data is too small and the distribution too uneven to use it for generating a model as many areas are not covered. In addition to the archaeological data also present day geographical data for the complete the study area is available. This data can be used to assess whether areas have been suitable for keeping sheep but of course a suitable environment alone is no indication that herders have been present there in prehistoric times.



Reference system: WGS 84; data: SRTM 500 (Jarvis et al., 2008).

Natural regions according to numbers:

**1** Pannonian basin, **2** Carpathian mountain range, **3** Pontic steppe, **4** Continental Pontic steppe, **5** Northern Balkan mountain range, **6** Lower Danube reach, **7** Central Balkan mountain range, **8** Peloponnes mountain range, **9** Northern mediterranean coast region, **10** Maramara region, **11** Aegean region, **12** Pontus region, **13** Central Anatolia, **14** Anatolian south coast, **15** Eastern Anatolia, **16** Azerbaijan, **17** Caucasus mountains, **18** Mesopotamia, **19** Syrian desert **20** Levant, **21** Zagros system, **22** Central Iranian Highlands, **23** Northern Iranian mountains.

☆ Tell Sabi Abyad: wool is assumed to have been produced here for the first time, around 6200 BC (Rooijakkers, 2012).

- Archaeological sites with bone finds (n = 401) (Küchelmann et al. in preparation).
- Archaeological sites with finds of spindle whorls (n = 23) (Grabundzija and Russo, 2016).

Figure 3.1: Map of the study area with excavation sites and borders of a partition into major landscape units [14].

To deal with the uncertain data we introduce an agent-based modeling approach that only utilizes the geographical data in the model construction and leaves the archaeological data for discussing the plausibility of the results. The basic idea for the model is the hypothesis that the woolly sheep has mainly been propagated through the prehistoric human society and that the mobility of the shepherds is mainly influenced by the search for a suitable environment. The simulation of the ABM then generates a distribution of the population in the physical space and a hypothetical realization of the historical spreading process which can be analyzed and compared with the archaeological data. From the microscopic dynamics of the ABM emerges a larger-scale spreading process between the major landscape units of the study area, which are also visualized in Figure 3.1. The ABM can be utilized to estimate exchange rates between these regions and through the application of the model reduction framework of Chapter 2 we can construct a metapopulation model for the spreading process on the mesoscale. The derivation of a PDMM from the ABM will, however, be more challenging in the real-world application than it has been for the conceptual model of the Guiding Example 6.

In this section, we will first discuss the ABM for the woolly sheep in detail and in comparison to [14, 15] the model is formulated such that it fits to the general ABM framework from Chapter 1. Afterwards we will go into the details of the model reduction to a PDMM and compare the results with the ABM simulations to assess the approximation quality.

### 3.1.1 ABM for the Spreading of the Woolly Sheep

In this subsection, we define the ABM for innovation spreading in ancient times, in particular for the spreading of the woolly sheep in the Near East and South-East Europe for the time period starting at 6200 BC. At first we will discuss the basic model assumptions for the agent system in a descriptive way. Afterwards, we use the formalism of Chapter 1 to define the corresponding formal ABM as a Markov process. Then, we take a look at the available real-world data for our application and how we can integrate it into the model. Finally, we discuss the simulation of the model and the analysis and interpretation of the results.

#### Model assumptions

The first step of our model construction is to clarify what entity an agent represents in the model. It would only be natural to assume that when we model the spreading of a new variant of a species, i.e., the mutation of the woolly sheep, that the agents or at least some of them should be the individual members of the species itself. In our case, however, we assume that in addition to the spreading driven by intraspecies dynamics, e.g., the migration of sheep due to the need for foraging, the interaction with humans has been a major factor. We even go so far that we assume that the woolly sheep as a domesticated species was mainly propagated due to human interactions and that we can neglect the allegedly significantly slower spreading that occurred without human influence. Thus, an agent in the model represents a group of herders that keep sheep in their flock and are thus able to adopt the woolly sheep. The woolly sheep itself is then to be viewed as an innovation that can be present or absent among the herders. The spreading of the woolly sheep is thus an example for the spreading of innovations in ancient human societies.

The agents are located within the area of interest for the study and characterized by a position on the map and the presence (or absence) of the innovation. To keep the model from being too complex we only explicitly model the mobility of the agents within the study area and their interactions. Instead of also defining rules for demographic changes of the population, we keep the number of agents constant.

For the mobility of the agents we assume that the main cause of migration for the herders is the need for a suitable environment for feeding their flock. In addition, we also assume that agents want to avoid conflicts over resources and thus keep a distance between each other that is, however, not too large such that the interaction with other agents is still possible.

While the exchange between groups of herders has certainly been more complex, we assume that the agents can interact with all other agents that are nearby. The propagation of the innovation is modeled as a transmission from one agent to another due to a pairwise interaction, i.e., a second order adoption.

## Model formulation

**State space:** To formalize our model assumptions using the framework of Chapter 1 we define the domain  $\mathbb{X} \subset \mathbb{R}^2$  as the map of our study area and define for the status of the agents the set  $\mathbb{S} := \{0, 1\}$ , with 0 indicating the absence and 1 indicating the presence of the innovation of the woolly sheep. We assume a constant number of  $n_a$  agents that are characterized by a position in the domain  $\mathbb{X}$  and an innovation status from  $\mathbb{S}$ . We denote again the agents' positions by a vector  $X$ , the status of all agents with a vector  $S$  and the system state by  $Y := (X, S)$ . The system state space is consequently defined as  $\mathbb{Y} := \mathbb{X}^{n_a} \times \mathbb{S}^{n_a}$ .

**Agent mobility:** As the main motivation for migration of an agent  $\alpha$  is the search for a suitable environment we define a *suitability landscape*  $V : \mathbb{X} \rightarrow \mathbb{R}$  that encodes for each location of the domain an assessment how fitting the local environment is for herding sheep, with lower values indicating a higher suitability. As mobility patterns for foraging often exhibit Lévy walk or Brownian motion characteristics [93] we assume that for the agents of our model a Brownian motion  $\mathbf{B}_\alpha(t)$  on  $\mathbb{X}$  is the driving force. The Brownian motion is on the one hand a good choice for modeling the exploratory nature of the foraging mobility and on the other hand also accounts for unknown reasons to change the location, e.g., as a reaction to a catastrophic event. The resulting process is then a diffusion process in the suitability landscape  $V$  which we already have studied before in this work for simpler settings. The metastability of this process embodies the return to suitable environments and thus the two main characteristics of foraging dynamics are represented in the model.

It remains now to include the influence of the positions of the other agents as an additional factor for the mobility process. A popular choice for modeling this kind of dynamics is through interatomic potentials such as the Morse potential [94] which has also been utilized to model swarming behavior [95]. Thus, for each agent  $\alpha$  we introduce an interaction potential  $U_\alpha : \mathbb{X}^n \rightarrow \mathbb{R}$  with

$$U_\alpha(\mathbf{X}(t)) = \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^n -c_A \exp\left(-\frac{\|\mathbf{x}_\alpha(t) - \mathbf{x}_\beta(t)\|}{l_A}\right) + c_R \exp\left(-\frac{\|\mathbf{x}_\alpha(t) - \mathbf{x}_\beta(t)\|}{l_R}\right), \quad (3.1)$$

where  $\|\cdot\|$  refers to the Euclidean distance,  $c_A \leq 0$  is the constant specifying the strength of the attraction force,  $c_R \leq 0$  is the corresponding constant for repulsion force strength and  $l_A, l_R \leq 0$  are the respective decay rate constants. As long as we choose  $c_R > c_A$  and  $l_a > l_R$  we have the repulsion from other agents dominating for short distances and the attraction forces dominating for longer distances. The potential  $U_\alpha$  has then a unique local minimum at an optimal distance where both attraction and repulsion forces are 0.

Combining these two influences we can formulate the mobility process for an agent

$\alpha$  of our model with the stochastic differential equation

$$d\mathbf{x}_\alpha(t) = -(\nabla V(\mathbf{x}_\alpha(t)) + \nabla U_\alpha(\mathbf{X}(t))) dt + \sigma dB_\alpha(t), \quad (3.2)$$

where  $\sigma > 0$  denotes a diffusion constant,  $V$  the suitability landscape and  $U_\alpha$  the interaction potential of agent  $\alpha$ . The mobility process for the complete system of agents  $\mathbf{X}(t)$  can be viewed as a diffusion process in a generalized potential landscape  $(V + U)$  [35] that is defined on the system state space. The suitability landscape  $V$  will be designed such that the border  $\partial\mathbb{X}$  of the domain is an unsuitable area and in addition we impose reflecting boundary conditions, i.e., a motion due to randomness that leads to an exit of the domain is rejected. Thus, the mobility process is restricted to the compact set  $\mathbb{X}^{n_a}$  and there exists a unique equilibrium distribution for the positions of the agent system.

**Innovation spreading:** For the spreading process between the agents we define a jump process and assume a second order adoption from innovation status 0 to innovation status 1 as defined in Equations (1.7) and (1.8). We define for each agent  $\alpha$  the status change vector  $v_{01}^{(\alpha)}$  as in Equation (1.5) that represents an adoption of the innovation and the corresponding second order adoption rate function

$$f_{01}^{(\alpha)}(X, S) = \delta_0(s_\alpha) \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{n_a} c_{01} \delta_1(s_\beta) d_r(x_\alpha, x_\beta), \quad (3.3)$$

with  $c_{01} > 0$  a rate constant and  $r > 0$  the interaction radius. We assume that there is no transition back to status 0, i.e., the agents do not dismiss the innovation once it was adopted. The jump process for the adoption dynamics is then given by

$$\mathbf{S}(t) = \mathbf{S}(0) + \sum_{\alpha=1}^{n_a} e_\alpha \mathcal{P}^{(\alpha)} \left( \int_0^t f_{01}^{(\alpha)}(\mathbf{Y}(s)) ds \right), \quad (3.4)$$

where  $e_\alpha$  is the  $\alpha$ -th  $n_a$ -dimensional unit vector and the  $\mathcal{P}^{(\alpha)}$  are independent unit rate Poisson processes.

**Combined process:** The combined process for the evolution of the agent system is then given by the stochastic differential equation

$$d\mathbf{Y}(t) = -(\mathbb{1}_{\mathbb{X}^{n_a}}, 0_{\mathbb{S}^{n_a}}) (\nabla(V + U)(\mathbf{Y}(t))) dt - \sigma d\mathbf{B}(t) \\ + \sum_{\alpha=1}^{n_a} v_{01}^{(\alpha)} f_{01}^{(\alpha)}(\mathbf{Y}(t)) d\mathcal{P}^{(\alpha)}(t),$$

with  $\mathbb{1}_{\mathbb{X}^{n_a}}$  denoting the 1-function on  $\mathbb{X}^{n_a}$ ,  $0_{\mathbb{S}^{n_a}}$  denoting the 0-vector on  $\mathbb{S}^{n_a}$  and  $\mathbf{B}(t)$  a Brownian motion in  $\mathbb{X}^{n_a}$ .

## Data integration

From the available data we use the present-day geographical data to assess whether an area of our domain has been suitable for herding sheep. The environmental data contains four geographical features (see Figure 3.2 a-d):

1. elevation data derived from the Shuttle Radar Topography Mission with resolution  $500\text{m}\times 500\text{m}$  (SRTM 500) [96]
2. topographic compound index (TCI) [97]
3. terrestrial landforms [98,99]
4. data on the soil texture [100]

The TCI represents the tendency of water to accumulate at any point in the catchment and the tendency for gravitational forces to move the water downslope. These factors influence the suitability of an area for grazing sheep and thus will be used to construct the suitability landscape for the mobility model. The environmental data covers the whole area of interest with a resolution of  $500\text{m}\times 500\text{m}$  and is assumed to be static for the simulated time period.

For the construction of the suitability landscape we can infer four main environmental factors that influence sheep-keeping from the four data sets: (1) elevation, (2) water availability (TCI), (3) different geomorphological landforms (such as ridges, flat plains) and (4) soil texture [14]. For each of these four environmental factors we have given an assessment which values correspond to a high suitability and thus we have four functions  $F_i : \mathbb{X} \rightarrow \mathbb{R}$ ,  $i = 1, \dots, 4$  on the domain that map each location to the evaluation of the suitability. The final suitability landscape is then simply constructed as  $V := \sum_{i=1}^4 F_i$  which corresponds after rescaling to an unweighted average of the environmental factors. A weighted average could be an alternative option for the construction, either with the help of additional expert knowledge or the utilization of an approximate bayesian computation approach as is described in [14] to estimate the weights.

A majority of the bone data is located in suitable areas of the landscape  $V$ , which can be interpreted that our assessment of the suitability is plausible. The presence of ovicaprid bones at a position, however, only indicates that a sheep or goat either died or was buried at that location and does not allow for the conclusion that the location of the finding also was suitable. Vice-versa the absence of bone findings does not allow for the conclusion that an area was unsuitable, especially in our situation of many missing data points. The assessment of the suitability by the geographers that have been involved in the project is thus a lot more convincing than an attempt to validate the suitability landscape with the limited archaeological data.

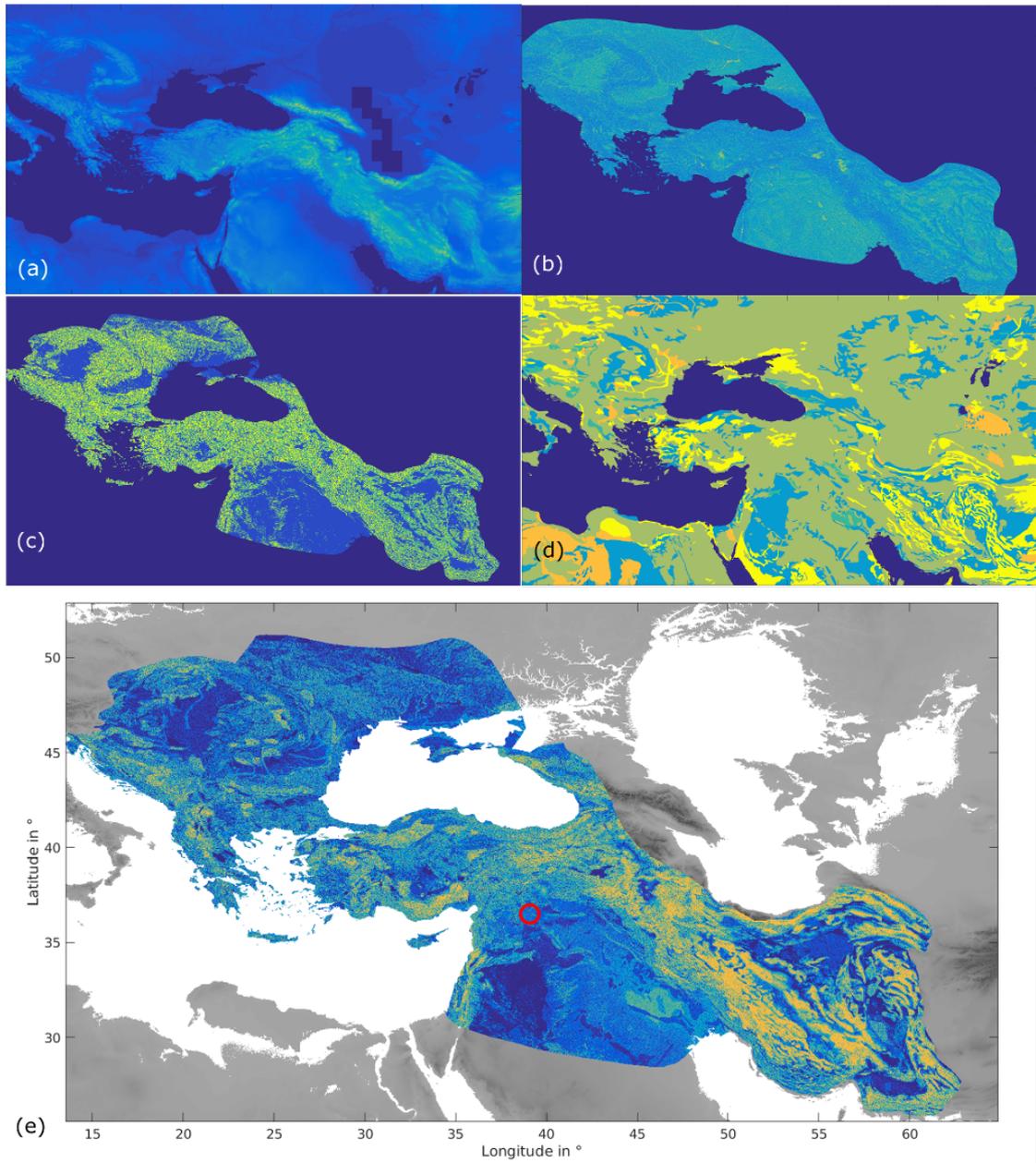


Figure 3.2: Visualisation of the different geographical features (a)-(d) that are used to construct the suitability landscape  $V$  and a visualisation of  $V$  itself (e). The colormap ranges from blue to yellow with blue indicating the most suitable and yellow the least suitable areas. The assumed origin of the woolly sheep in northern Syria is marked with a red circle in (e).

## Simulation details

We choose the number of agents to be  $n_a = 4000$  which corresponds to a population of roughly 40000-100000 herders within the study area if we consider an agent to represent a small group of 10-30 people. While not much is known about the population density in that time frame simulations for the population density in the area of today's Greece [101] suggest that the chosen number of herders is reasonable. We measure the time of our model in simulated years and choose the default time step size for the simulations to be  $\Delta t = \frac{1}{365}$  which corresponds to a simulated day. The suitability landscape  $V$  and the diffusion coefficient  $\sigma$  are scaled such that the average distance that an agent moves within a time step is roughly 1km. The constants and rates for the interaction potential  $U$  are scaled such that we have the situation of short range repulsion and long range attraction and an optimal distance between agents of approximately 1km.

For the parameters of the second order adoption rate functions we consider two different scenarios. The first scenario is the one discussed in [15] with the adoption rate constant being  $c_{01} = 8$  (which corresponds to roughly 45 days of pairwise interaction until an adoption happens on average) and an interaction radius of  $r = 10\text{km}$ . As this parameter choice for the adoption rate constant is rather high we have the situation that the adoption process is faster than the mixing of agents within suitable areas. This has an influence on the approximation quality of a model reduction to a PDMM. Thus, it is interesting to also consider a second scenario for comparison where we have in addition to the time scale separation of the mixing of the agents within and the transition of the agents between suitable areas of the landscape also a time scale separation between the adoption process of the agents and the mixing within the metastable sets of the suitability landscape. Therefore, we choose  $c_{01} = 0.05$  and  $r = 50\text{km}$  as parameter values for the second scenario. As these choices are vastly different than the parameters of the first scenario the question remains whether the choices are still reasonable and allow for an interpretation of the results in the archaeological context. Considering the parameter choices for modeling hunter-gatherer societies in the next section I would argue that the parameter values of the second scenario of this model are still plausible, even though they are mainly motivated by the applicability of the model reduction framework.

We initialize the agent system with a uniform distribution of positions and only simulate the mobility process until the positions are approximately distributed according to the equilibrium distribution of the diffusion process within the suitability landscape. Then we initialize the spreading process by setting the status of all agents that are close to the assumed origin of the woolly sheep to 1. The time is then initialized as 0 (which corresponds to 6200 BC). As some agents are isolated and not reachable, e.g., when positioned on an island, the simulation is stopped when 95% of the agents have adopted the innovation. In both scenarios, we have the situation of rather sparsely distributed agents that have only few connections, which results in a generally low total adoption rate. Also by the construction of our model at most 3999 adoption events can occur within more than a million time steps. Thus, it is clear that for the simulation of the

model the event-based approach of Algorithm 5 is more accurate and more efficient than the alternatives. The algorithm has not to be further adapted for the model, however, as for both the evaluation of the adoption rate functions and the evaluation of the interaction potential we need a neighborhood computation, it is clear that the neighbor lists need to be stored for the time step duration such that the computation needs to be only done once in each time step. The neighborhood computations as well as the evaluation of the adoption rate functions and the interaction potential represent the majority of the computational effort and unfortunately scale superlinearly with the agent numbers. As the agent system of the model is already quite large compared to the conceptual models of the examples in Chapter 1 this is reflected in a comparably long simulation time. For a single realization of the ABM a node equipped with the hardware available at the time of writing needs about 30 hours of computation time.

## Results and analysis

As the model is too complex for a formal analysis we will only do a descriptive analysis of the simulation results. Let us begin with discussing the example simulation of Scenario 1 that is visualized in Figure 3.3. In the first snapshot, we see the initial system state of the time when the spreading process starts. In the second snapshot, we can observe that the spreading within the Fertile Crescent region happens very fast until the agents encounter the first geographical barriers given by unsuitable areas. While during the next 1000 simulated years the woolly sheep has already spread to most regions in the Near east the transition to Europe only occurs after a longer waiting time of an additional 900 simulation years (see snapshots 3 and 4). After this critical transition the spreading within Southeastern Europe happens again on a faster time scale and after roughly 2800 simulated years the simulation ends with today's Ukraine being the last region of our study area that is reached (see snapshots 5 and 6).

There are many metastable system states that occur whenever a geographical barrier of unsuitable terrain needs to be passed by an agent to spread the innovation to the next suitable area and thus the metastability in the spreading process is inherited from the metastability of the mobility process. The greatest time scale difference is induced by the transition from the Near East to Europe over the Bosphorus. This metastable behavior is also observable when looking at the macroscopic spreading process, i.e., the number of woolly sheep adopters over time, for several simulations as well as the average over a sample of 42 ABM simulations for Scenario 1 and 50 ABM simulations for Scenario 2 (see Figure 3.4). The spreading speed of the average simulation slows down significantly once the first transitions through unsuitable terrain need to occur to continue the spreading. For Scenario 2 we can observe that the adoption speed of the average trajectory also slows down noticeably again for the metastable transition into Europe. In the single simulations, we can observe many time periods throughout which the number of adopters over time is almost constant, especially during the long waiting time for the transition to Europe.

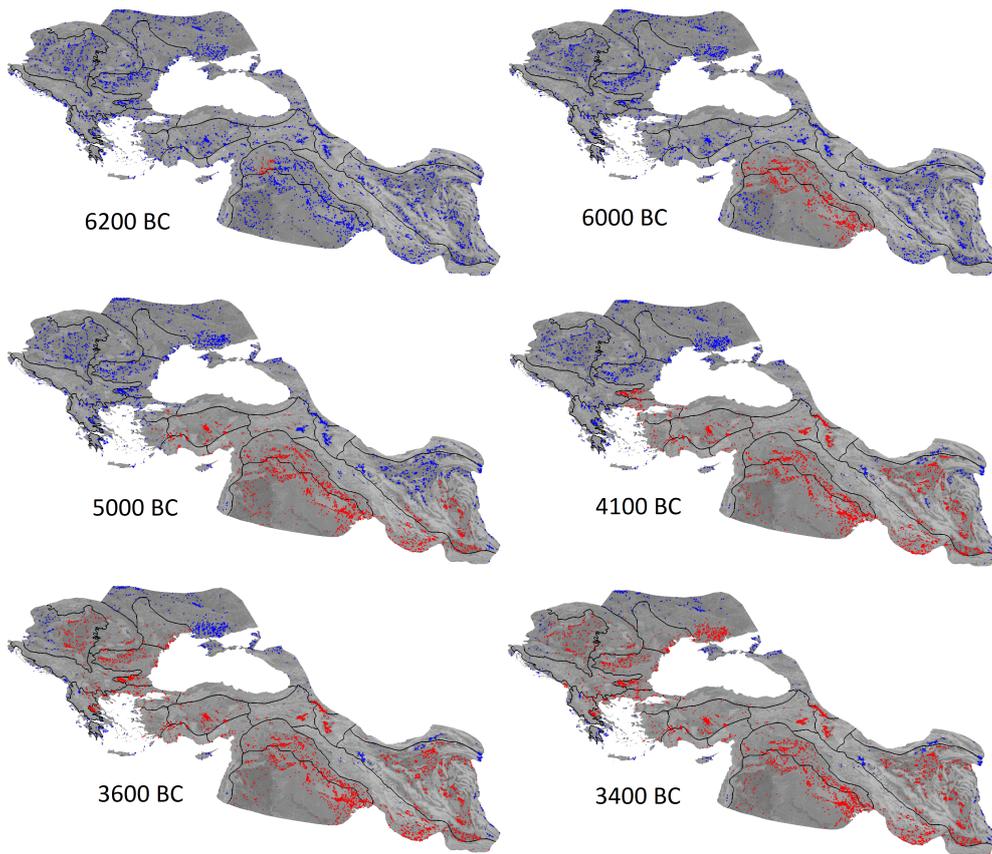


Figure 3.3: Visualization of six snapshots of an ABM simulation for the spreading of the woolly sheep. Blue dots mark agents without the innovation and red dots agents that have adopted the woolly sheep. The black borders indicate a partition of the domain into 23 major geographical landscape units. [14]

Also when considering the mean first hitting time of the 23 specified geographical regions for agents with status 1 the metastability of the mesoscopic spreading process between the regions is visible (see Figure 3.5). The transition to Europe via the Bosphorus happens on a slower time scale than the spreading within the Near East and within South-East Europe. The much greater interaction radius of Scenario 2 makes it more likely to transmit the innovation over geographical barriers without agents fully crossing them leading to significant differences for the mean first hitting time of some mountainous regions. For a more detailed look at the first hitting time distributions, especially for the critical transition to Europe, the sample size is unfortunately too small to get meaningful results.

The observed differences between the two scenarios raise the question how sensitive is the model to parameter changes? It is clear that higher adoption rates lead to a faster spreading, but only up to the point where each contact between agents with a different

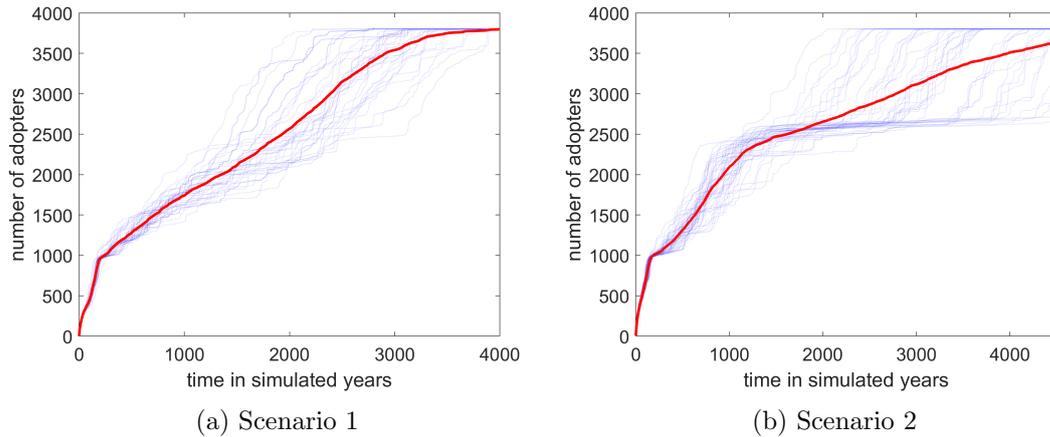
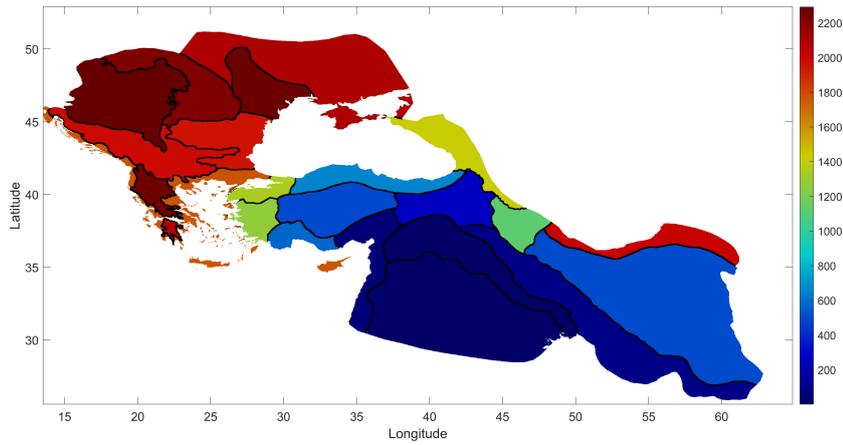


Figure 3.4: Number of adopters over time for the ABM simulations (blue curves) and the average over all trajectories (red curve) for each scenario. The variance increases greatly with time and for the trajectories of the single simulations metastability in the adoption process can be observed. In comparison, the macroscopic spreading process of Scenario 2 is more metastable and multiple time scale separations can be observed in the average trajectory.

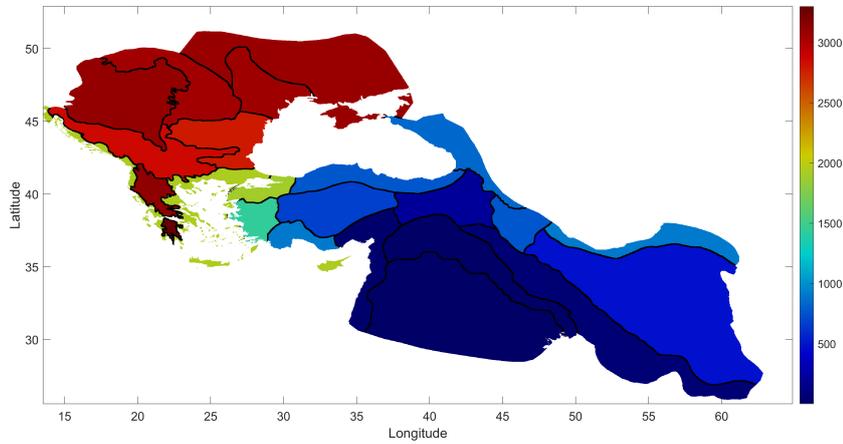
status is expected to lead to an instantaneous adoption event. In Scenario 1, we do not yet have this situation, but the high adoption rate already leads to differences in the observed macroscopic patterns. The spatial spreading path is, however, unaffected by changes to the adoption rate and primarily influenced by our choice for the suitability landscape and the scaling of the diffusion process. An increase in the number of agents  $n_a$  will lead to an increased frequency of transitions between suitable areas and thus also speed up the overall spreading process. The mobility patterns might also slightly change due to the interaction potential. Finally, an increased interaction radius  $r$  leads to an increase in the average number of connections between agents and as we have observed can also affect the spreading path if chosen too large. A sensitivity analysis that was performed for Scenario 1, however, suggests that the model is robust to small changes to the interaction radius  $r$  [15].

### Validation and interpretation

One can only speculate whether or not the true prehistorical spreading path is correlated with the outcome of our simulation. The main reason we can not validate our model is that the available archaeological data is very sparse and obscure. The data suitable for validation consists of the spindle whorls from 23 sites that are located only in the Panonian Basin [102]. Also, we can not generate new data to reproduce the true dynamics in order to validate our results. For all these reasons a model validation in the common sense of natural science appears impossible to us. The simulated spreading path can be understood as a spreading path hypothesis which has to be discussed and



(a) Scenario 1



(b) Scenario 2

Figure 3.5: Mean first hitting times for the 23 major geographical landscape units measured in simulated years for both scenarios.

evaluated with the help of additional expert knowledge [15].

The mean first hitting time of the Panonian Basin for agents with status 1 is a few hundred years later than the dating of the oldest spindle whorls of the data set. This can be explained with the model probably overestimating the critical transition time to Europe. As the model does not allow for sea travel and we only assume movements of the foraging type but not the use of established routes the Bosphorus is probably a too strong bottleneck.

### 3.1.2 PDMM Approximation

While the small number of ABM simulations already provides a nice illustration and possibly some insights for archaeologists about the hypothetical spreading path based on the assumptions about the mobility patterns in ancient times, a more thorough analysis and estimation of critical transition time distributions is infeasible because of the small sample size. As the sample size can only be increased with a lot of additional computation time it already takes too much effort to even generate a sufficiently large sample size for one scenario, let alone for multiple parameter sets. Thus, we are interested in reducing the simulation effort by applying model reduction techniques.

While the macroscopic spreading dynamics could probably also be learned by a method like SINDy [103] with use of the right basis functions, as it has been successfully applied to other ABMs [104], the mesoscopic patterns that we are especially interested in would be lost. The model reduction framework of Chapter 2 fits much better to our situation, as only the microscopic details of the ABM will be lost through a PDMM approximation, but the spatial patterns on the mesoscale can be preserved. Thus, we will now discuss the construction of a PDMM based on the ABM for innovation spreading in ancient times.

#### Identification of subpopulations

The first step of the PDMM construction is the definition of the subpopulations. As we already have a partition into 23 geographical regions given this would be the easiest choice for choosing the subpopulations. However, as we can observe in the microscopic trajectory some regions contain multiple agent clusters that could be considered metastable and some agent clusters are positioned at the borders between the partition sets. It is thus clear that the given partition into 23 major landscape units is not metastable, so we can not expect that the resulting model would have a good approximation quality. So, we have to begin with identifying the metastable sets with respect to the mobility process.

As we do not have an analytical expression for the transfer operator of the mobility model we can not directly apply the spectral approach to metastability. While we could utilize a square root approximation [105] to estimate a transfer operator for the diffusion in the suitability landscape, this approach would neglect the influence of the interaction potential. For the generalized potential landscape that includes all influences of the mobility model, however, we can not apply this technique as the state space is too high dimensional. We thus choose to apply an approach based on clustering methods.

The most commonly used method is probably the kmeans algorithm [106], which performs well for centroid shaped clusters but can fail with correctly identifying oddly shaped clusters. In this situation, density based clustering methods perform very well and a current state of the art method in this regard additionally utilizes an hierarchical

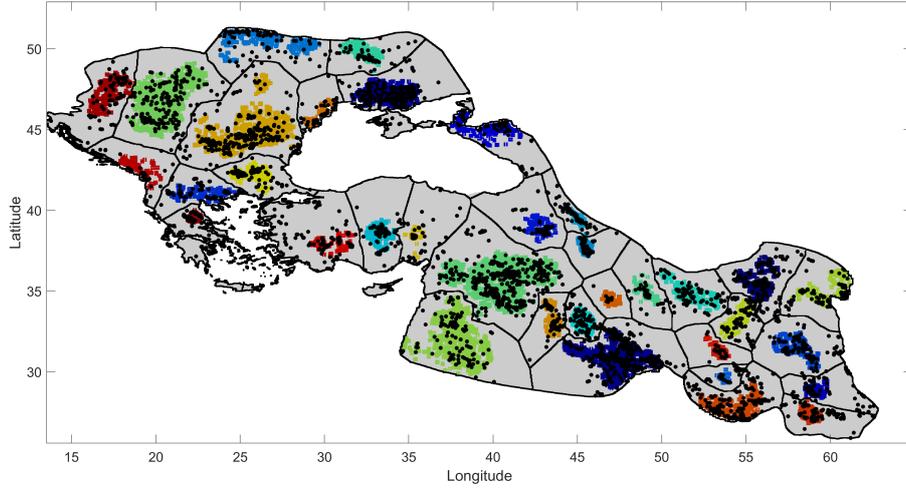


Figure 3.6: Visualization of agent clusters that have been identified with HDBSCAN and a corresponding full partition that is constructed by a Voronoi tessellation based on the core sets. The core sets are marked by colored regions, the borders of the full partition set are indicated by black lines and for comparison the positions of an ABM snapshot are visualized by black dots.

clustering approach [107], which improves the performance for varying data point densities. Thus, the hierarchical density based clustering method HDBSCAN [107] will be our method of choice to identify the metastable areas of our domain with respect to the mobility process.

We apply the HDBSCAN algorithm to the position data sampled from ten different realizations of the ABM for a reasonable minimum cluster size and get as a result a set of  $m = 35$  spatial clusters (see Figure 3.6). Depending on the approach we have now various possibilities to define the subpopulations of the PDMM. The first possibility would be to utilize the core set approach and directly define the subpopulations  $C_1, \dots, C_m$  as the clusters from the HDBSCAN results. The second possibility would be a full partition approach, where we define a partition  $A_1, \dots, A_m$  of the domain by utilizing a Voronoi tessellation based on the clusters (see Figure 3.6). In the following, we will consider both approaches for constructing the model and discuss the differences for the estimated transition rates and interaction probabilities. There are of course also possibilities in between the two approaches, e.g., a core set approach with enlarged core sets containing the clusters but not yet forming a full partition. The enlargement of the core sets in a Voronoi fashion would lead to monotonically increasing transition rates for the PDMM, with the model based on the full partition being the one with the maximal transition rates.

### PDMM definition

We define the system state of the PDMM approximation by an  $m \times 2$  matrix  $N$  which is initialized by drawing an initial distribution of population members according to an estimated invariant measure  $\hat{\mu}$  for the mobility process. The members of the subpopulation that includes the origin of the woolly sheep are initialized with status 1 and for all other subpopulations the members are initialized with status 0. We define the spatial transitions between the subpopulations as a jump process generated by a rate matrix  $\Lambda$  and as the type of dynamics is the same we define the deterministic dynamics of the internal adoption process as we did for the Guiding Example 2.2.5, i.e., we define for each subpopulation

$$\hat{f}_{01}^{(k)}(N) = c_{01} b_k N_0^{(k)} N_1^{(k)}, \quad (3.5)$$

with  $b_k$  being the interaction probability within the subpopulation  $k$ . However, instead of using a forward Euler scheme to integrate the adoption rate functions to propagate the internal adoptions, we directly use the analytical solution given by the logistic function, i.e., we have for consecutive jump times  $t_0 < t_1$

$$\mathbf{N}_1^{(k)}(t) = n_0^{(k)} \left( 1 + e^{-c_{01} b_k n_0^{(k)} t} \left( n_0^{(k)} - \mathbf{N}_1^{(k)}(t_0) \right) \right)^{-1},$$

with  $n_0^{(k)} := \mathbf{N}_0^{(k)}(t) + \mathbf{N}_1^{(k)}(t)$ . This way we can use an adapted stochastic simulation algorithm, where we only need to draw the jump events for the spatial transitions and thus have an additional gain in computational efficiency.

### Parameter estimation from ABM data

We need to estimate in total three different parameters for each subpopulation  $k$ : the invariant measure  $\hat{\mu}_k$  for the initialization of the model, the  $k$ th row of the rate matrix  $\Lambda$  and the interaction probability  $b_k$ . For the estimation of the rate matrix  $\Lambda$  we use the maximum likelihood estimator from Section 2.1.3. We can estimate the invariant measure of subpopulation  $k$  by calculating the average number of agents assigned to subpopulation  $k$  denoted by  $\hat{n}^{(k)}$  and setting  $\hat{\mu}_k := \frac{\hat{n}^{(k)}}{n_a}$ . For the core set approach the assignment to the subpopulations is realized by calculating a milestone trajectory from the simulation data. Finally, we estimate  $b_k$  by calculating the interaction probability for agents assigned to subpopulation  $k$  for each simulation snapshot and averaging over all snapshots.

As the core set approach leads in general to better estimates of the transition rate matrix  $\Lambda$  we first define the subpopulations of the PDMM as  $C_1, \dots, C_m$  and estimate the rate matrix for the core set MSM of the mobility model  $\Lambda(C)$ , as well as the interaction probabilities  $b_k(C)$  and the invariant measure  $\hat{\mu}(C)$ . Unfortunately the resulting model is not working as intended, as the transition rates from core sets located in the Near East to core sets located in Southeastern Europe are all estimated as 0, i.e., there are no spatial transitions possible from the Near East to Europe. While at first this result may seem devastating we actually gain the insight, that the original mobility model

is not able to actually perform transitions over the Bosphorus. The reason for that is unclear, but it can be presumed that because of the narrow landscape the boundary conditions are often imposed and agents tend to get stuck at the Bosphorus. On the other hand, because of the interaction potential the stuck agents repel other agents that are entering the area of the Bosphorus. Nevertheless, the transmission of the innovation is still possible over the Bosphorus through a chain of adoption events.

Since the PDMM approximation failed for the core set approach, the next step is to apply the same procedure for the full partition case and define the subpopulations of the PDMM as  $A_1, \dots, A_m$ . We estimate the rate matrix for the core set MSM of the mobility model  $\Lambda(A)$ , as well as the interaction probabilities  $b_k(A)$  and the invariant measure  $\hat{\mu}(A)$ . The interaction probabilities and the invariant measure are estimated almost the same as for the core set approach, but as expected the estimated rate matrix  $\Lambda(A)$  consists of much higher transition rates. While the resulting model is working as intended the transitions between the subpopulations occur much faster than in the ABM and thus also the simulated spreading process and the estimated first hitting times are too fast. This can be explained by the frequent recrossings of the borders of the partition sets and is a common problem for full partition MSMs, as we also pointed out in Section 2.1.3.

While both approaches led to inaccurate models each of the two PDMMs had some desirable properties. While the core set PDMM was overall not working the transition rates between the subpopulations located within one of the two continental regions have been estimated closer to the ABM. On the other hand, while too fast the full partition PDMM produced a working model with a network of subpopulations only consisting of one connected component. So it is natural to consider a weighted average of the two models. This way the two components of the core set PDMM will be connected due to the inclusion of the positive full partition transition rates between the subpopulations that include the Bosphorus region. On the other hand the full partition transition rates that have been too high get reduced due to the weighted averaging with the core set MSM rate matrix.

To be precise, for the combined PDMM approach we define the subpopulations as the full partition sets  $A_1, \dots, A_m$ , we choose a weight constant  $w \in [0, 1]$  and estimate the parameters for the model as  $\Lambda(w) := w\Lambda(C) + (1-w)\Lambda(A)$ ,  $b_k(w) := wb_k(C) + (1-w)b_k(A)$  and  $\hat{\mu}(w) := w\hat{\mu}(C) + (1-w)\hat{\mu}(A)$ . And indeed, the resulting model produces acceptable results which we can compare to the ABM results in the next section to assess the approximation quality. In the machine learning community the concept of averaging different models is also used, e.g., for neural networks [108], and the resulting averaged models can have an increased approximation quality.

### 3.1.3 Comparison of both Approaches

We will now look at the results of the PDMM approximation in comparison with the respective results of the ABM for innovation spreading in ancient times. For both sce-

narios we are able to generate a much larger sample of PDMM simulations because of the reduced computational effort. The simulation of a single PDMM simulation takes on average only a few seconds of computation time with the same hardware used for the ABM simulations, which represents a massive gain in efficiency of 5 orders of magnitude! This means in addition to visualizing the average macroscopic spreading and the expected arrival times between the regions on the mesoscale we are able to actually sample the distributions of the critical transition times, especially for the critical transition to Europe, or even spreading path statistics.

### Scenario 1

For Scenario 1 we choose the weight  $w = 0.6$  for the averaging of the model parameters, which means that the model parameters of the combined PDMM are closer to those derived from the core set approach.

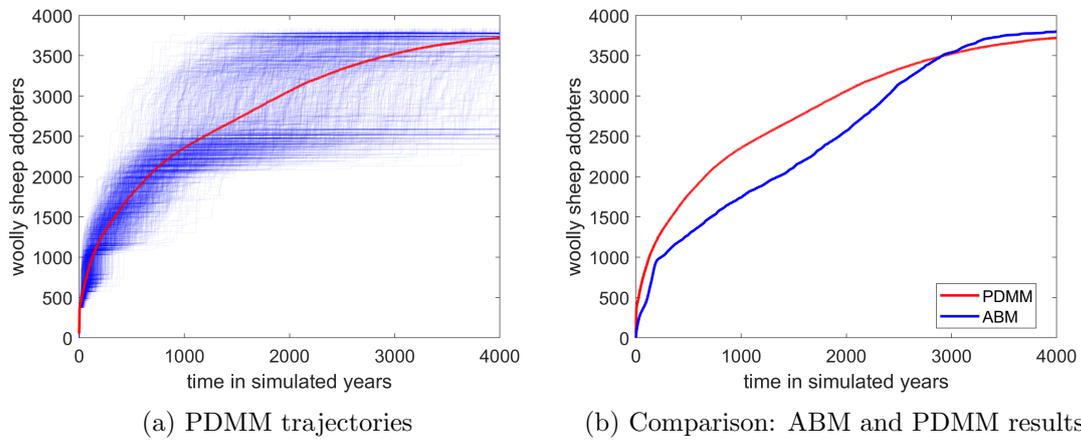
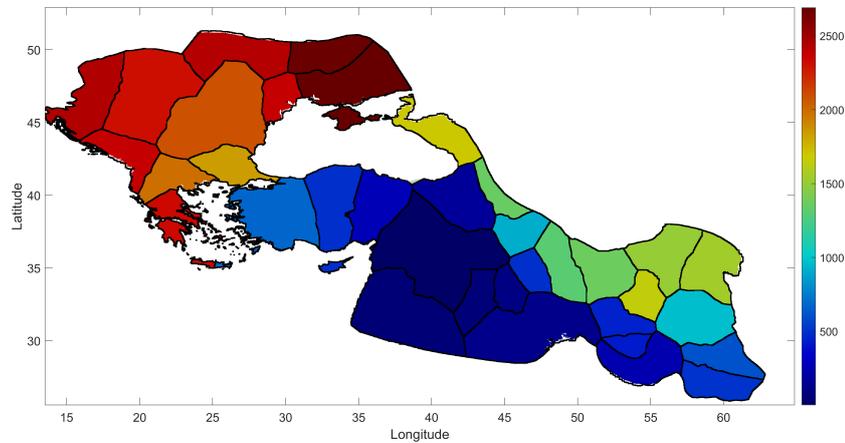


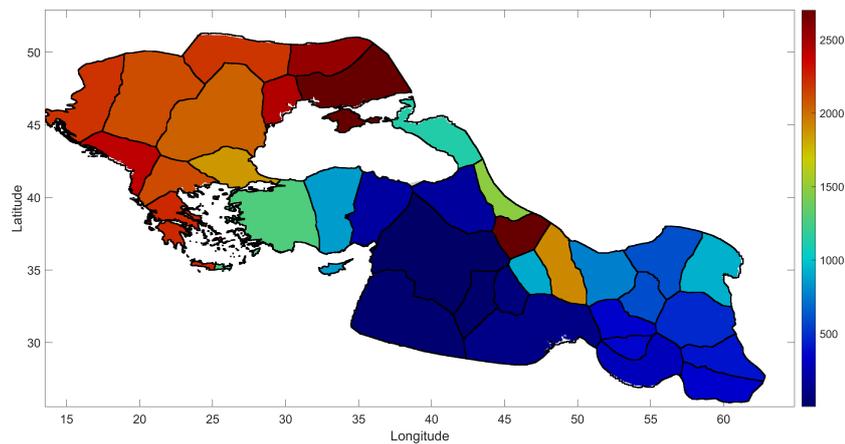
Figure 3.7: PDMM results for Scenario 1. In (b) the averaged macroscopic trajectories for the number of woolly sheep adopters of both model types are compared.

In Figure 3.7 (a) the trajectories of 1000 PDMM simulations are visualized as well as the average over the sample. We can observe the increasing variance over time and the metastability of the macroscopic trajectory for the number of woolly sheep adopters. Compared to the ABM simulations we get a much better picture about the variation of the trajectories due to the increased sample size. Comparing the average trajectories of the PDMM and the ABM we can observe that there is some deviation and that the PDMM spreading is faster. It is unclear, whether an increased sample size would also lead to a smaller error, but we know that the parameter choice for the adoption rate constant  $c_{01}$  is so high that the adoption process is faster than the mixing within the subpopulations, which can be a source of error and cause a reduced approximation quality. We can see that even in the beginning of the simulation, before the main source of error due to the approximation of the spatial transition takes effect, we already have

a significant difference in the average macroscopic spreading paths.



(a) ABM for Scenario 1



(b) PDMM for Scenario 1

Figure 3.8: Mean first hitting times for the partition sets of the full partition used for the PDMM construction. The results for the ABM have been averaged over 42 simulations while for the PDMM a sample of 1000 simulations was used.

On the mesoscale we can overall observe a similar pattern for the first arrival of the woolly sheep (see Figure 3.8), though there are clearly some sets with larger deviations for the mean first hitting time. The slower transition over the Bosphorus is well captured in the reduced model and the estimated mean first arrival of the woolly sheep in Europe is very similar in both models (ABM: after 1834 simulation years, PDMM: after 1860 simulation years.)

## Scenario 2

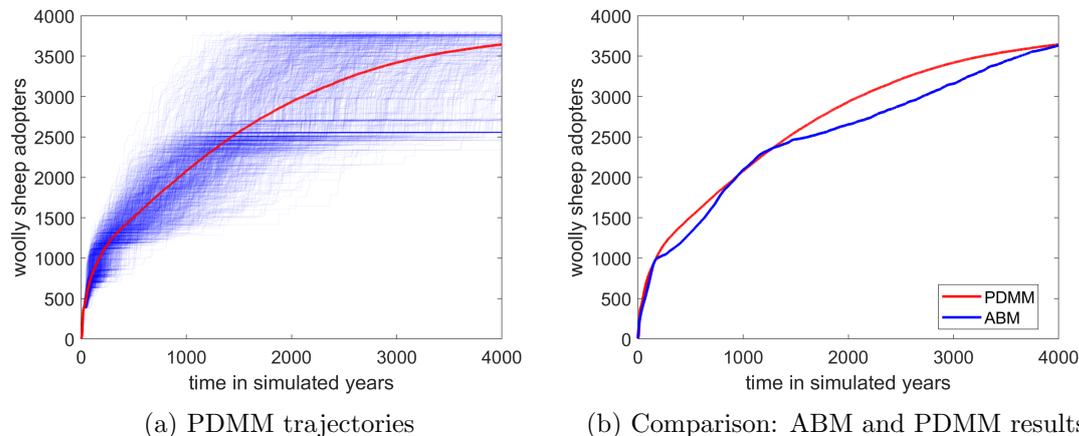
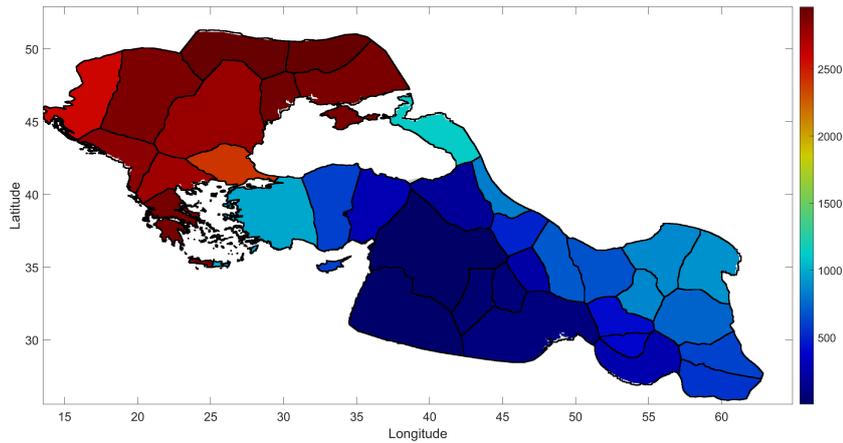


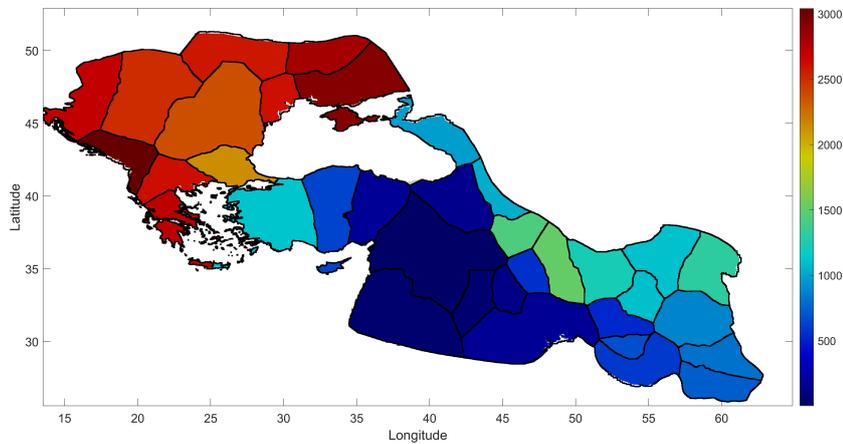
Figure 3.9: PDMM results for Scenario 2. In (b) the averaged macroscopic trajectories for the number of woolly sheep adopters of both model types are compared.

For Scenario 2 we choose the weight  $w = 0.66$  for the averaging of the model parameters, which means that the model parameters of the combined PDMM are again closer to those derived from the core set approach. In Figure 3.9 (a) the trajectories of 1000 PDMM simulations are visualized as well as the average over the sample. We can observe the increasing variance over time and the metastability of the macroscopic trajectory for the number of woolly sheep adopters. Compared to Scenario 1 the characteristics of the PDMM and the ABM trajectories are much more similar. While the ABM trajectories for Scenario 1 have been less metastable (at least for the small sample) the PDMM trajectories had almost instantaneous spreading of the innovation within the subpopulations. In comparison, the trajectories of the PDMM for Scenario 2 feature a slower and thus smoother spreading within the subpopulations as is also characteristic for the ABM trajectories. The fit of the average trajectories (see Figure 3.9 (b)) is much better than it has been for Scenario 1 and we can reasonably assume that an increase in size for the ABM sample would lead to an even better fit of the curves. As we have constructed Scenario 2 to have faster mixing within the subpopulations compared to the speed of the adoption process this potential source of error is no longer present. Indeed, especially in the beginning of the simulation we now also have a very good fit.

On the mesoscale we can again observe overall a similar pattern for the first arrival of the woolly sheep (see Figure 3.10) and compared to the first scenario there are no sets with extremely large deviations of the mean first hitting time. The slower transition over the Bosphorus is again well captured in the reduced model but the estimated mean first arrival of the woolly sheep in Europe is this time significantly faster in the PDMM (ABM: after 2401 simulation years, PDMM: after 2132 simulation years). It is interesting to note that in the southwestern part of the domain the PDMM simulation actually



(a) ABM for Scenario 2



(b) PDMM for Scenario 2

Figure 3.10: Mean first hitting times for the partition sets of the full partition used for the PDMM construction. The results for the ABM have been averaged over 50 simulations while for the PDMM a sample of 1000 simulations was used.

produces larger first hitting times compared to the ABM. This can be explained by the large interaction radius of 50km in Scenario 2, which can lead to a spreading over geographical barriers.

As in the PDMM the interactions between agents of different subpopulations are neglected this effect is not present, as there has to actually happen a spatial transition between the subpopulations to enable the spreading. For the PDMM the increased interaction radius only induces an increased interaction probability  $b_k(w)$ . A similar situation might also appear for a few subpopulations that are surrounded by mountains

in the situation of Scenario 1.

All comparisons between the models have to be taken with a grain of salt as the sample size for the ABM simulations is too small in both cases. Some of the observed differences or similarities between the model outcomes might only be a result of the specific sample of ABM simulations and thus simply caused by the stochastic nature of the model.

### Critical transition times

Finally, we briefly take a look at the distribution of the critical transition time of the woolly sheep to Europe estimated from a sample of 10000 simulations for each scenario (see Figure 3.11).

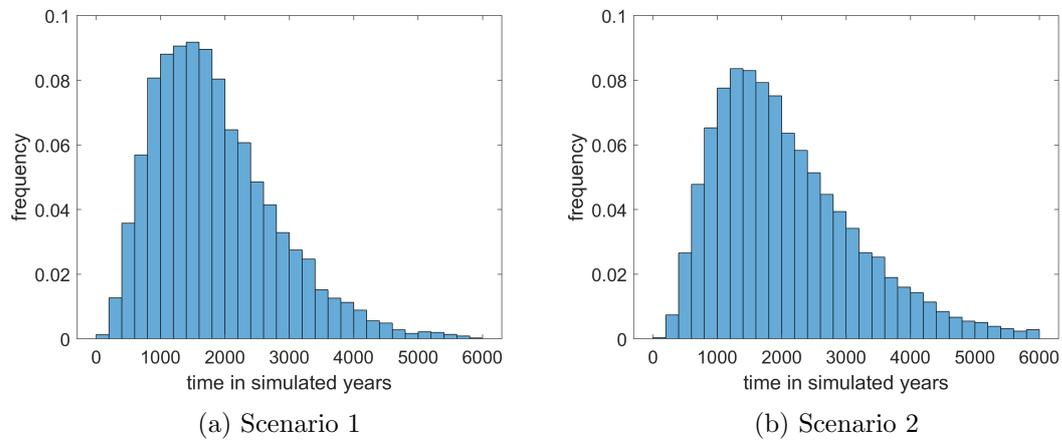


Figure 3.11: PDMM results for the distributions of the critical transition time to Europe for both scenarios.

Overall the shape of the distributions is heavy tailed and similar for both scenarios. While a direct comparison with the ABM data is not meaningful because of the difference in sample size, a Bayesian statistics approach [109, 110] might be suitable to assess the likelihood of the critical transition time sample from the ABM simulations being drawn from the distribution that is estimated by the PDMM. The application of Bayesian methods might even be used to find an optimal weight  $w$  for defining the combined model parameters. This, however, will be beyond the scope of this work and be left as a possible topic for further research.

## 3.2 Cultural Evolution of Hunter-Gatherer Societies

The second application of our modeling framework will again be the situation of a pre-historical process. This time we are interested in developing a model for the cultural evolution of hunter-gatherer societies in Central Africa from 120000 BP until present time. While in the modeling scenario from the previous section we were designing the model for innovation spreading with a specific prehistoric process in mind, in this section the intention is to design a more general model that can be easily adapted to various more concrete applications.

Central Africa homes some of the largest, most resilient and genetically ancient hunter-gatherer populations in the world [111, 112] spanning 120,000 years. It is relatively safe to assume that until the advent of the Neolithic human populations were exclusively composed of hunter-gatherer bands [113]. Thus gaining insights about the mobility of hunter-gatherers is of importance not only for cultural evolution but for many prehistoric processes. A key element of hunter-gatherer mobility is the adaptation to changing environments [114–117] and the changes in mobility patterns are assumed to be a major influence for population and cultural dynamics [118–120]. The evolution of complex culture has parallels to the evolutionary dynamics of the genetic inheritance system [121–123] and a key question is the role of interconnectedness of populations [124–127]. While this question has already been extensively studied [125, 128–130] there is still a need for formal and explicit models to investigate the influence of environmental changes to the interconnectedness of hunter-gatherer populations.

We will develop in this section a spatio-temporally explicit agent-based model to examine the socio-ecological drivers of hunter-gatherer demographic and mobility patterns and how these patterns could have affected the process of cultural evolution. While the model will be designed utilizing real-world data from hunter-gatherers and the environment of Central Africa it can easily be adapted to model the mobility of hunter-gatherers from other regions of the world. The model for cultural evolution will be rather general and conceptual such that it can be adapted to many more concrete research questions. Still, we are able to explore which aspects of hunter-gatherer social structures are required for the emergence and maintenance of complex, cumulative culture and we can offer insights into the adaptive nature of a foraging lifestyle. This model is also presented in [17] with a greater focus on the case study and the interpretation of the results in the anthropological context.

### 3.2.1 ABM for Mobility and Cultural Evolution of Hunter-Gatherers

In this subsection, we formulate an agent-based model for the cultural evolution of hunter-gatherer societies. We will at first state the modeling assumptions in a descriptive way and then formalize our assumptions to define the Markov jump process for the ABM. We will use the experience that we gained through the modeling of innovation spreading in ancient times for the spreading of the woolly sheep and build upon the model from the last section. The model will be adapted to include the different mobility patterns of hunter-gatherers in Central Africa and expanded such that we can formulate the more complex dynamics of the cultural evolution process.

#### Model assumptions

The agents are located in the Central African region including the Congo Basin. As we are interested in the interactions between hunter-gatherer societies on this rather large spatial scale we choose our agents to represent camps of hunter-gatherers instead of individual human beings. The agents are characterized by the camp location on the map, the number of camp members and a cultural status. The mobility of the agents thus does not describe the mobility of the camp members but represents the residential mobility of the hunter-gatherer camps, which is one of the defining features of hunter-gatherers around the world and has been proposed to have important implications for cultural transmission and evolution [114, 115, 127, 131–133].

We assume that the residential mobility is mainly motivated by the need for a suitable environment that provides enough resources for foraging within the vicinity of the camp location [115]. The agents are thus searching for a suitable environment, but also want to avoid conflicts over resources with other agents, i.e., overlapping foraging areas. The speed of the movements is influenced by the environmental conditions, i.e., the friction of the landscape.

The camp populations are subject to demographic changes that depend on the carrying capacity of the environment at the camp location. A lack of resources leads to a decline of the population, while a sufficient supply leads to population growth. The population of a camp cannot grow indefinitely, as after a certain point it can become too large to stay organized as a single unit [134, 135]. In this case, there is a fission into two smaller camps. Given that hunter-gatherer survival and reproductive abilities depend on camp-wide division of labour, cooperation and sharing [131, 135–137], we also assume that an agent needs a minimum number of camp members to be able to survive. An agent that does not meet this required number either faces extinction or can join another nearby camp.

The culture of an agent consists of some number of attributes, referred to as cultural *features*, each of which can take a number of values (or *traits*). Changes to the culture of an agent happen on the one hand due to intrinsic processes and on the other hand due to

pairwise interactions with other camps that can lead to a cultural transmission. For the interactions we assume a dependence on the distance of the camps, i.e., agents that are closer to each other interact more often. We assume that there is a dynamical distinction between cultural features that tend to evolve towards greater complexity [138], e.g., technology, and cultural features that are not subject to ecological pressures and where complexity is not necessarily associated with greater payoffs, such as songs, folktales or any other stylistic traits [139, 140]. We assume different intrinsic processes for the two cases and a bias in the transmission dynamics in the first case such that more complex traits are favoured.

## Model formulation

We formulate our model on the micro-scale as an agent-based model similar to the approach that we utilized for modeling innovation spreading in ancient times. The basic setting considers a set of  $n_a$  agents that follow rules for their spatial movement and the social interactions that govern cultural transmission. While at first we assume the number  $n_a$  to be static we will move on to  $n_a$  being rather an upper bound for the number of agents and explain how we realize a changing number of active agents.

An agent  $\alpha$  represents a hunter-gatherer camp that at every point in time  $t$  has: (1) a *position*  $x_\alpha(t)$ , representing a location of the camp in terms of latitude and longitude coordinates; (2) a *camp population*  $d_\alpha(t)$ , representing the number of people who live in the camp and (3) a *cultural status*  $s^{(\alpha)}(t)$ , representing a cultural tradition or technology of the camp. Thus, the state of the agent  $\alpha$  at time  $t$  is given by  $y_\alpha(t) := (x_\alpha(t), d_\alpha(t), s^{(\alpha)}(t))$  and  $Y(t) := (X(t), D(t), S(t))$  denotes the system state at time  $t$ . Formally, a position is denoted by a continuous state variable taking values in the compact domain  $\mathbb{X} \subset \mathbb{R}^2$ , a camp population taking values in  $\mathbb{R}$  and a cultural status is a discrete state variable that takes values in the countable set  $\mathbb{S}$ , such that  $Y(t) \in \mathbb{Y} := \mathbb{X}^{n_a} \times \mathbb{R}^{n_a} \times \mathbb{S}^{n_a}$ .

There are three processes that govern the change of the system state of the ABM. The first is the (residential) *mobility* of the agents that we model as a diffusion process in a suitability landscape  $V$ , where agents change their position in  $\mathbb{X}$  according to the suitability and friction of their physical environment and in reaction to the movements of the other agents.

The second process is the *demographic* change of the population vector  $D(t)$  which is modeled deterministically with population growth and decline depending on the carrying capacity of the landscape as well as the system state  $Y(t)$ .

Finally, we have the evolutionary *cultural dynamics* for which we use a common adapted version of Axelrod's definition [141] in which culture is defined to be a set of attributes that are subject to social influence [142, 143]. We consider the culture of an agent to be composed of a finite number of  $c$  cultural features, each represented by an integer

value. A feature could be knowledge about a certain kind of tool or technique but also a shared belief or societal value (see Section. Thus, in a general setting we consider as status space  $\mathbb{S} \subset \mathbb{N}^c$ .

Next, we will define the three processes of our ABM for the mobility and cultural dispersal of hunter-gatherer societies in more detail. Although the model and analytical tools presented in the following sections are designed to be generalizable across temporal, environmental and ethnographic settings, we detail the present parametrization our model with data from contemporary hunter-gatherer populations living in Central Africa.

### Mathematical modeling of agents' mobility

In our model, the movement of the agents, i.e. the relocation of camp positions, is governed by the environmental influences, interaction with other agents and possible unknown influences. Formally, these dynamics are generated by a stochastic differential equation, such that the movement of every agent  $\alpha$  is given by

$$d\mathbf{x}_\alpha(t) = -\nabla (V(\mathbf{x}_\alpha(t), t) + U_\alpha(\mathbf{X}(t))) dt + \sigma(\mathbf{x}_\alpha(t)) dB_\alpha(t), \quad (3.6)$$

where  $V : \mathbb{X} \times \mathbb{T} \rightarrow \mathbb{R}$  is the time-dependent suitability landscape of the environment,  $U_\alpha : \mathbb{X}^{n_a} \rightarrow \mathbb{R}$  the interaction force between agents,  $\sigma : \mathbb{X} \rightarrow \mathbb{R}$  the friction-dependent scaling function for the noise and  $B_\alpha(t)$  a standard Brownian motion in  $\mathbb{R}^2$ . We now proceed with more details on the different components of the equation (3.6).

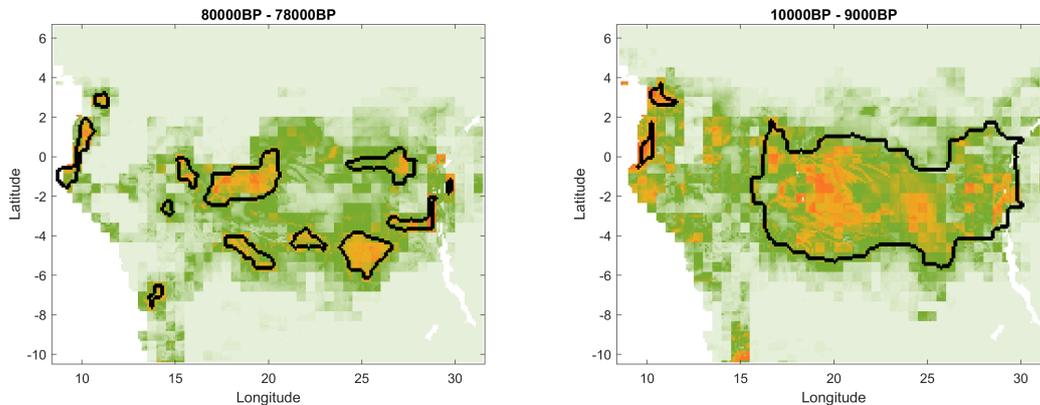


Figure 3.12: Two snapshots of the time-dependent suitability landscapes  $V$  at 80,000BP and 10,000BP with borders of mobility clusters marked in black.

**Environmental influence:** We account for the possible environmental factors by constructing a time-dependent suitability landscape  $V : \mathbb{X} \times \mathbb{T} \rightarrow \mathbb{R}$  that determines which areas of  $\mathbb{X}$  are attractive for the agents [15] (see Figure 3.12 for a visualization). For the construction of the suitability landscape  $V(\cdot, t)$  we utilize a bioclimatic environmental niche model (ENM) from which we derive the likelihood of a hunter-gatherer camp being

present for each position in the map [111]. Higher suitability values from the ENM in a particular area correspond to a higher attractiveness of that area for hunter-gatherers.

**Social interaction:** The mobility of agents is also governed by the position of other agents, such that agents at long distances are attracted towards each other and agents at short distances are repelled from each other. Intuitively, these attraction-repulsion forces between agents represent the trade-off between avoiding isolation and benefiting from social interaction for material or cultural exchanges and avoiding conflicts over territories or scarce resources [114, 115, 117, 144]. Thus, for each agent  $\alpha$  we define an interaction potential  $U_\alpha : \mathbb{X}^{n_\alpha} \rightarrow \mathbb{R}$  with

$$U_\alpha(X(t)) = \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^{n_\alpha} -c_A \exp\left(-\frac{\|x_\alpha(t) - x_\beta(t)\|}{l_A}\right) + c_R \exp\left(-\frac{\|x_\alpha(t) - x_\beta(t)\|}{l_R}\right), \quad (3.7)$$

where  $\|\cdot\|$  refers to the Euclidean distance,  $c_A \leq 0$  is the attraction potential constant,  $c_R \leq 0$  is the repulsion potential constant and  $l_A, l_R \leq 0$  are the respective decay rate constants [94]. This is the same kind of interaction potential that we used for the modeling of innovation spreading in Section 3.1.1 and the relations of the constants are chosen in a similar fashion such that there is again an optimal distance between the agents. The dominating term is the short range repulsion, which leads to agents maintaining a minimum distance from each other so that the foraging areas (i.e., areas regularly used for subsistence activities) of each camp do not overlap. This assumption has a solid foundation in the literature about hunter-gatherers [114, 127, 144].

**Stochastic effects:** The stochastic part of the SDE in (3.6) is represented by a scaled Brownian motion, and it prevents the system from becoming stationary even if every agent has found a position with high suitability and with enough distance from other agents. The characteristics of diffusion within a suitability landscape captures the nature of the foraging mobility of hunter-gatherers, which is in this regard similar to the model in Section 3.1.1. The scaling of the Brownian motion determines how fast agents can travel, and hence it is dependent on the friction of the terrain. We define the scaling function  $\sigma : \mathbb{X} \rightarrow \mathbb{R}$  such that higher friction of the terrain implies a lower scaling of the Brownian motion. For environmental barriers, e.g., a cliff or a mountain, the scaling function will take values close to 0. In this way, it is very unlikely that an agent in our model moves through impassable terrain.

### Cultural evolution of agents

We encode the status of all agents in a vector  $S(t)$  with the entry  $s_i^{(\alpha)}(t)$  being the trait value of agent  $\alpha$  for feature  $i$  at time  $t$ . The cultural evolution of agents is modeled as a Markov jump process where the possible jumps between different trait values of a feature are encoded by the status change vectors  $v_j, j \in \{1, \dots, J\}$  with  $J \in \mathbb{N}$ . For two consecutive event times  $t_1 < t_2$  and the jump event at time  $t_2$  corresponding to vector  $v_j$ , we can write the change to the system state due to status changes as  $(Y(t_2)) = Y(t_1) + v_j$

with  $v_j$  only acting on the status vector  $S$  and not on the other state variables. Each of the  $J$  cultural status changes has a time-dependent rate specified by an associated adoption rate function that depends on the evolution of the agent system.

For the status changes we consider two different event types: Events that result from intrinsic processes of camps, e.g., the development of a new innovation or loss of knowledge, by first order status changes and events that result from exchange or social learning between camps, e.g., the transmission of a trait from one camp to another, by second order adoptions. Transmissions between two agents are governed by the agents' spatial proximity that is captured in the structure of a time-dependant *interaction network* that we utilize to define the second order adoption rate functions.

Additionally, we consider two types of cultural features: *Progressive* and *Non-progressive*. The former represent, e.g., tools or technological innovations, where efficiency and pay-offs increase with an increasing number of component elements [138]. As the knowledge about more different or more efficient variants of a tool is beneficial we bias the transmission dynamics to favor the transfer of knowledge [145, 146]. On the other hand, "Non-progressive" cultural features represent those cultural domains not subject to ecological pressures and where complexity is not necessarily associated with greater efficiency [139, 140]. In this case the transmission dynamics will be unbiased.

### Interaction network

We assume that for the interaction of two agents spatial proximity is required and construct a network from the agents' positions  $X(t)$ , where two distinct agents  $\alpha$  and  $\beta$  are adjacent at time  $t$  if their Euclidean distance is closer than a specified interaction radius  $r$ , i.e.,  $\|x_\alpha(t) - x_\beta(t)\| \leq r$ . In our model we assume more frequent interactions, i.e., a higher interaction rate  $\varphi_1$ , in the closer neighborhood of the agents compared to long distance interactions that are less frequent with rate  $\varphi_2$ . We thus choose an interaction radius  $r_1 > 0$  for defining the short range and another interaction radius  $r_2 > r_1$  for defining the long range interactions between agents.

Then, we define the adjacency matrix  $A(Y(t))$  for the time-dependant weighted interaction network between agents by

$$A_{\alpha\beta}(Y(t)) := \begin{cases} \varphi_1, & \text{if } \|x_\alpha(t) - x_\beta(t)\| \leq r_1 \\ \varphi_2, & \text{if } r_1 < \|x_\alpha(t) - x_\beta(t)\| \leq r_2 \\ 0, & \text{else} \end{cases} \quad (3.8)$$

with  $\varphi_1, r_1$  corresponding to short and  $\varphi_2, r_2$  corresponding to long range interactions. In Figure 3.13, we plot a part of the interaction network from one simulation at time 80000 BP.

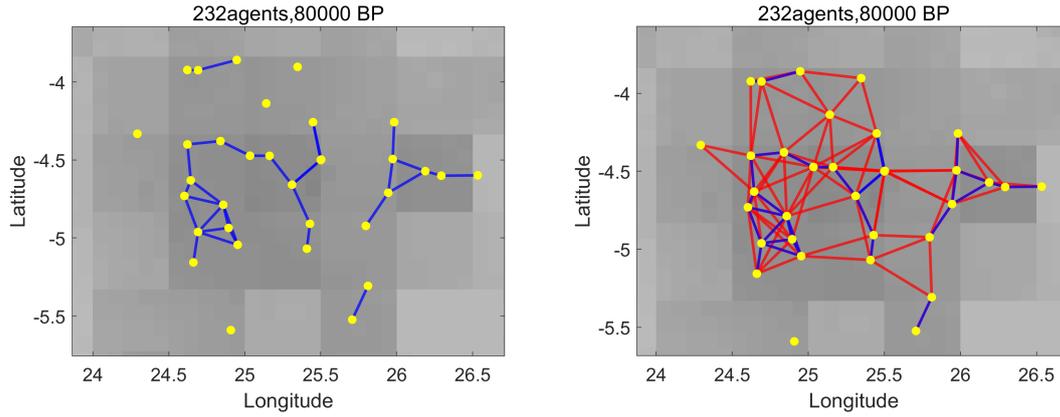


Figure 3.13: Snapshot of a simulation for illustrating the short range interactions (left) and the weighted interaction network (right). Blue edges correspond to edge weights for short range and red edges to edge weights for long range interactions.

### Progressive cultural features

In the case of progressive features we consider only gradual changes of the trait values, i.e., depending on the type of event an agent either increases or decreases its trait value by 1. We thus define for a progressive feature  $i$  and each agent  $\alpha$ , the  $n_a$ -dimensional vector  $e_i^{(\alpha)}$  which has a value 1 for feature  $i$  and agent  $\alpha$  and value 0 otherwise. Combining the status change vector  $v_i^{(\alpha)} := e_i^{(\alpha)}$  with the 0-vector on  $\mathbb{X}^{n_a} \times \mathbb{R}^{n_a}$  gives us then the associated state change vector  $\tilde{v}_i^{(\alpha)}$  for the ABM system state. We consider the set containing all  $v_i^{(\alpha)}$  and  $-v_i^{(\alpha)}$  as the set of possible status change events with  $2cn_a$  elements.

**First order status changes:** We choose constant rates  $\gamma_i, \lambda_i > 0$  for gain and loss of information in feature  $i$  respectively and define the corresponding status change rate functions by setting

$$g_i^{(\alpha)}(Y(t)) := \gamma_i$$

for the status change vector  $v_i^{(\alpha)}$  and

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

for the status change vector  $-v_i^{(\alpha)}$ . We can write down the jump process for the first order status changes as

$$\begin{aligned} & \sum_{\alpha=1}^{n_a} \sum_{k=1}^c \mathcal{Q}_k^{(\alpha)} \left( \int_0^t g_k^{(\alpha)}(Y(\tau)) d\tau \right) v_k^{(\alpha)} - \sum_{\alpha=1}^{n_a} \sum_{k=1}^c \mathcal{R}_k^{(\alpha)} \left( \int_0^t l_k^{(\alpha)}(Y(\tau)) d\tau \right) v_k^{(\alpha)} \\ &= \sum_{\alpha=1}^{n_a} \sum_{k=1}^c \mathcal{Q}_k^{(\alpha)}(\gamma_k t) v_k^{(\alpha)} - \sum_{\alpha=1}^{n_a} \sum_{k=1}^c \mathcal{R}_k^{(\alpha)} \left( \int_0^t l_k^{(\alpha)}(Y(\tau)) d\tau \right) v_k^{(\alpha)} \end{aligned}$$

where  $\mathcal{Q}_k^{(\alpha)}, \mathcal{R}_k^{(\alpha)}$  are independent unit rate Poisson processes.

**Second order status changes:** We then define the status change rate function for the gain of knowledge in feature  $i$  by transmission of information to agent  $\alpha$  by

$$f_i^{(\alpha)}(Y(t)) := \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^n \max \left\{ 0, s_i^{(\beta)}(t) - s_i^{(\alpha)}(t) \right\} \cdot A_{\alpha\beta}(Y(t)). \quad (3.9)$$

We assume that different variants of a cultural feature are adopted independently of each other and this implies that a greater difference of the trait values of two agents is leading to a higher likelihood that an interaction between the two agents will induce a gain of knowledge in that feature for the agent with the lower trait value. This is analogous to people learning from more knowledgeable or skilled individuals [147, 148]. An agent can only increase the trait value through interaction if it has a contact with an agent that has a higher trait value in that feature. As long as there are neighbors with higher trait values there is a positive rate for an adoption event leading to an increment of the trait value.

We can write down the jump process for the second order adoptions as

$$\sum_{\alpha=1}^{n_a} \sum_{k=1}^c \mathcal{P}_k^{(\alpha)} \left( \int_0^t f_k^{(\alpha)}(Y(\tau)) d\tau \right) v_k^{(\alpha)}$$

where  $\mathcal{P}_k^{(\alpha)}$  are independent unit rate Poisson processes. Combining the processes for first and second order status changes we get as Markov jump process for the cultural evolution

$$\begin{aligned} \mathbf{S}(t) &= \mathbf{S}(0) + \sum_{\alpha=1}^{n_a} \sum_{k=1}^c \mathcal{P}_k^{(\alpha)} \left( \int_0^t f_k^{(\alpha)}(Y(\tau)) d\tau \right) v_k^{(\alpha)} \\ &\quad + \sum_{\alpha=1}^{n_a} \sum_{k=1}^c \mathcal{Q}_k^{(\alpha)}(\gamma_k t) v_k^{(\alpha)} - \sum_{\alpha=1}^{n_a} \sum_{k=1}^c \mathcal{R}_k^{(\alpha)} \left( \int_0^t l_k^{(\alpha)}(Y(\tau)) d\tau \right) v_k^{(\alpha)}. \end{aligned}$$

To illustrate the rules for the dynamics of progressive cultural features, in Figure 3.14 we plot two consecutive snapshots of one simulation run around 110000 BP. Here, we

consider a case of  $c = 3$  features and color the agents according to their cultural status, such that each value  $s_i^{(\alpha)}$ ,  $i = 1, 2, 3$  denotes one component of the RGB color vector. Thus, in the resulting plot agents of a similar status are colored by a similar color.

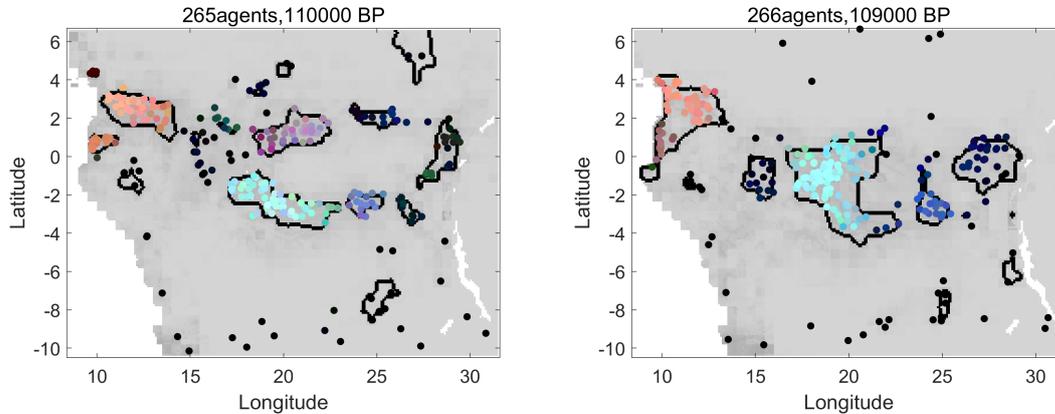


Figure 3.14: Two consecutive snapshots of a simulation run with  $c = 3$  progressive cultural features. Agents are depicted as dots and colored according to their cultural status. The borders of spatially grouped agents are marked black.

### Non-progressive cultural features

We consider for a non-progressive feature  $i$  a finite set  $\mathbb{S}_i$  of  $n_s$  traits. We assume the  $n_s$  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. We thus define for all trait values  $k \neq l \in \mathbb{S}_i$  and each agent  $\alpha$  a status change vector  $v_{ikl}^{(\alpha)} := (-k + l)e_i^{(\alpha)}$  encoding the switch from trait  $k$  to  $l$  in feature  $i$  for agent  $\alpha$  and the associated system state change vector  $\tilde{v}_{ikl}^{(\alpha)}$ .

**First order status changes:** We choose a constant rate  $\gamma_i$  for spontaneous changes of trait values in feature  $i$  (in analogy to evolutionary models also sometimes called mutations) and define for all possible status change vectors  $v_{ikl}^{(\alpha)}$  the corresponding rate functions by

$$g_{ikl}^{(\alpha)}(Y(t)) := \gamma_i \delta_k \left( s_i^{(\alpha)}(t) \right)$$

with  $\delta_k$  being the discrete indicator function of trait  $k$ .

**Second order status changes:** Similar to the progressive case we assume that interactions of agents require spatial proximity and thus consider the same weighted network defined by the adjacency matrix  $A(Y(t))$  in Equation (3.8). The dynamics on the interaction network however differ and are in the non-progressive case such that an agent copies the trait of a neighboring agent similar to models for opinion dynamics [141]. We define the rate function for the status change  $v_{ikl}^{(\alpha)}$  due to transmission of trait  $l$  to agent  $\alpha$  as

$$f_{ikl}^{(\alpha)}(Y(t)) := \sum_{\substack{\beta=1 \\ \beta \neq \alpha}}^n \delta_k(s_i^{(\alpha)}(t)) \delta_l(s_i^{(\beta)}(t)) \cdot A_{\alpha\beta}(Y(t)), \quad (3.10)$$

where  $\delta_k$  and  $\delta_l$  denote the discrete indicator functions of traits  $k$  and  $l$ . As long as there are neighbors with a different trait value in some feature, there is a positive rate for such a status change event in which an agent copies the trait value of a neighbor. The non-progressive cultural dynamics are illustrated in Figure 3.15 for a case of  $c = 3$  features that can again be visualized by RGB color vectors. Also in this case agents with a similar cultural status have a similar color in the plot.

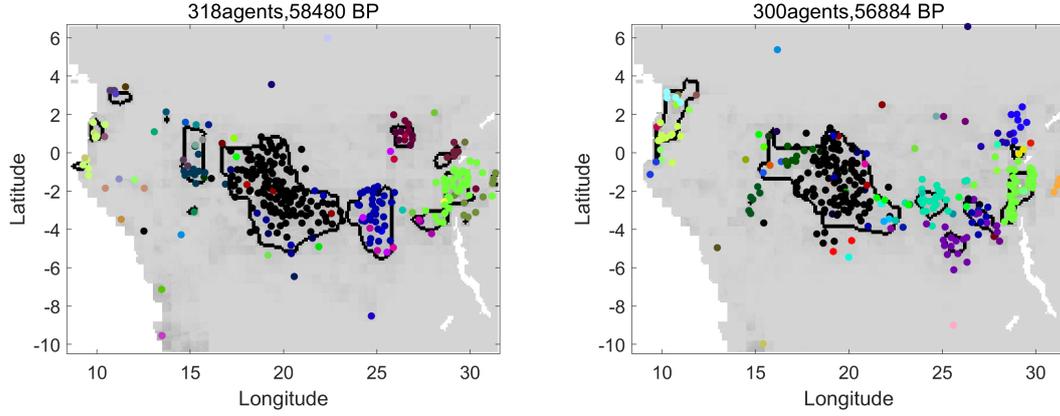


Figure 3.15: Two consecutive snapshots of a simulation run with  $c = 3$  non-progressive cultural features. Agents are depicted as dots and colored according to their cultural status. The borders of spatially grouped agents are marked black.

We can write down the jump process for the first order status changes as

$$\sum_{\alpha=1}^{n_a} \sum_{i=1}^c \sum_{k=1}^{n_s} \sum_{\substack{l=1 \\ l \neq k}}^{n_s} \mathcal{Q}_{ikl}^{(\alpha)} \left( \int_0^t g_{ikl}^{(\alpha)}(Y(\tau)) d\tau \right) v_{ikl}^{(\alpha)}$$

where the  $\mathcal{Q}_{ikl}^{(\alpha)}$  are independent unit rate Poisson processes and the jump process for the second order status changes as

$$\sum_{\alpha=1}^{n_a} \sum_{i=1}^c \sum_{k=1}^{n_s} \sum_{\substack{l=1 \\ l \neq k}}^{n_s} \mathcal{P}_{ikl}^{(\alpha)} \left( \int_0^t f_{ikl}^{(\alpha)}(Y(\tau)) d\tau \right) v_{ikl}^{(\alpha)}$$

where  $\mathcal{P}_{ikl}^{(\alpha)}$  are independent unit rate Poisson processes.

Combining the processes for first and second order status changes we get as Markov jump process for the non-progressive cultural evolution

$$\begin{aligned} \mathbf{S}(t) = \mathbf{S}(0) &+ \sum_{\alpha=1}^{n_a} \sum_{i=1}^c \sum_{k=1}^{n_s} \sum_{\substack{l=1 \\ l \neq k}}^{n_s} \mathcal{P}_{ikl}^{(\alpha)} \left( \int_0^t f_{ikl}^{(\alpha)}(Y(\tau)) d\tau \right) v_{ikl}^{(\alpha)} \\ &+ \sum_{\alpha=1}^{n_a} \sum_{i=1}^c \sum_{k=1}^{n_s} \sum_{\substack{l=1 \\ l \neq k}}^{n_s} \mathcal{Q}_{ikl}^{(\alpha)} \left( \int_0^t g_{ikl}^{(\alpha)}(Y(\tau)) d\tau \right) v_{ikl}^{(\alpha)}. \end{aligned}$$

## Demographics

For the demographics of the agents we model growth and decline of the population vector  $D(t)$  as deterministic processes that depend on the local carrying capacity of the environment. The local carrying capacity  $K_t(x)$  is the maximum number of people that can be supplied by foraging within the short range interaction radius  $r_1$  around the location  $x \in \mathbb{X}$  and depends linearly on the suitability  $V_t(x)$  (see [111] for empirical evidence of this relationship). In the case of two camps being close enough to have overlapping foraging areas, they each contribute to the local population of the other camp proportionally to the size of the intersection of their foraging areas. We thus define for an agent  $\alpha$  the *local population* at time  $t$

$$k_\alpha(t) := \sum_{\beta=1}^n \phi_{r_1}^{(\alpha\beta)} d_\beta(t)$$

with  $d_\beta(t)$  referring to the entries of  $D(t)$ . The sum of the population of all agents is weighted by the intersection proportion  $\phi_{r_1}^{(\alpha\beta)}$  of their foraging areas. In the case of  $\alpha = \beta$  we have  $\phi_{r_1}^{(\alpha\beta)} = 1$ , so a camp always contributes fully to its local population. If there is no intersection of the foraging areas around the camps  $\alpha$  and  $\beta$  we have  $\phi_{r_1}^{(\alpha\beta)} = 0$  and the population of camp  $\beta$  does not contribute to the local population of camp  $\alpha$ .

While the local population is smaller than the local carrying capacity, i.e.,  $k_\alpha(t) < K_t(x_\alpha)$ , the population  $d_\alpha(t)$  of a camp grows with rate  $\rho_g$ . If  $k_\alpha(t) > K_t(x_\alpha)$  we assume that  $d_\alpha(t)$  declines with rate  $\rho_d$ .

The population of a camp cannot grow indefinitely, as after a certain point it can become too large to stay organized as a single unit [134, 135]. We thus define a fission threshold  $h_{\text{fis}}$  that sets a maximum for the population size  $d_\alpha(t)$  of an agent. If a fission event occurs, i.e.,  $d_\alpha(t) > h_{\text{fis}}$ , the camp represented by agent  $\alpha$  splits up into two camps resulting in two new agents  $\beta_1$  and  $\beta_2$  being created with position  $x_{\beta_1}(t) = x_{\beta_2}(t) = x_\alpha(t)$  and  $s_i^{(\beta_1)}(t) = s_i^{(\beta_2)}(t) = s_i^{(\alpha)}(t)$  for all cultural features  $i$ . We assume that in the splitting process the members of the old camp are distributed equally between the new camps and thus set  $d_{\beta_1}(t) = d_{\beta_2}(t) = \frac{d_\alpha(t)}{2}$ . After the splitting process the agent  $\alpha$  is removed from the system.

Given that hunter-gatherer survival and reproductive abilities depend on camp-wide division of labour, cooperation and sharing [131, 135–137], we also assume that a camp needs a minimum number of members to be able to survive and thus define a fusion threshold  $h_{\text{fus}}$  that sets a minimum population for an agent. If  $d_\alpha(t) < h_{\text{fus}}$  the members of camp  $\alpha$  try to get taken in by another nearby camp within their interaction radius  $r_2$ . In the case that there is another camp  $\beta$  within  $r_2$  the two camps can merge, which is realized by modifying the agent  $\beta$  by increasing  $d_\beta(t)$  by  $d_\alpha(t)$  and adjusting the status of agent  $\beta$  in the progressive case. We set  $s_i^{(\beta)}(t)$  to the maximum value of  $s_i^{(\alpha)}(t)$  and  $s_i^{(\beta)}(t)$  for each feature  $i$ . When there are no nearby camps for a fusion process, then camp  $\alpha$  goes extinct. In both cases the agent  $\alpha$  is removed from the system.

The upper bound for the total population is given by the total carrying capacity  $\kappa_t := \int_{\mathbb{X}} K_t(x) dx$  at time  $t$ . This total carrying capacity at time  $t$  also induces a maximum number agents  $\eta_t$  at each time  $t$ . We choose the number of agents for our system to be  $n_a = \max_t \{\eta_t\}$  and realize the birth-death process for fusion and fission by distinguishing between *active* agents with  $d_\alpha > 0$  and *inactive* agents with  $y_\alpha = 0$ . This way we can realize the "birth" of an agent  $\alpha$  by assigning an inactive agent  $\alpha$  to a non-zero state when a fission event happens and the "death" of an agent  $\alpha$  by assigning it a zero vector state in case of a fusion or extinction event. Through our choice for  $n_a$  we ensure that we can always find inactive agents in case of birth events to assign to an active state. Finally, the mobility and status dynamics can be adjusted such that the change of inactive agents always is zero. The explicit definitions of the mobility and status dynamics can in this sense be understood as the rules governing the active agents of the system.

In Figure 3.16 we provide a compact overview of the model dynamics. The actual implementation of the model however does not exactly follow the flowchart but is realized again by utilizing the event-based approach of Algorithm 5 from Chapter 1. At the end of each time step there is an additional check for fission and fusion events that are realized as described in this section.

## Model overview

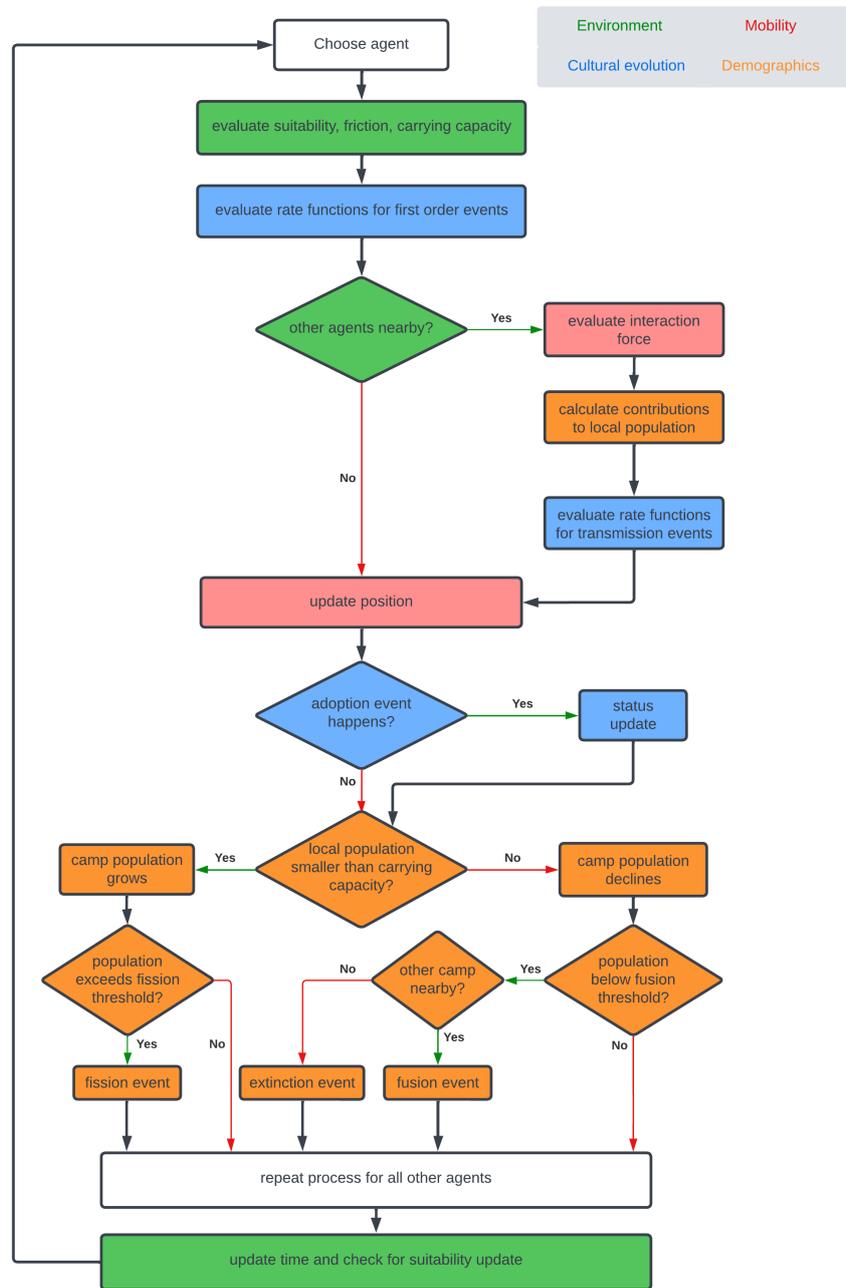


Figure 3.16: Flowchart of the model dynamics. The chart was created in Lucidchart ([www.lucidchart.com](http://www.lucidchart.com)).

## Model parametrization

The environmental niche model for generating the suitability landscapes was built using the data of in total 749 contemporary hunter-gatherer camps in Central Africa. The suitability value of a location on the map can be interpreted as a likelihood that the location is occupied by a contemporary hunter gatherer camp. The ENM then weights different environmental factors, e.g., the average temperature, to explain the observed likelihood for the present day data. By using a bias-corrected time series of global terrestrial climate and vegetation and the ENM we are able to estimate the suitability of the landscape up to 120000 years in the past [111, 149]. The time series of the palaeoclimatic reconstructions is given by time slices covering 1000-2000 years. The estimated suitability landscape is static during these time frames and only changes at the discrete time points where the time slices are interchanged. So, during the static periods of the suitability landscape the mobility process of the ABM can converge to an equilibrium state (if the convergence speed is significantly faster than the frequency of perturbations, which is the case for our model) and thus, even though the mobility dynamics are not reversible, we can analyze the metastability of the system and identify spatial agent clusters. This will be discussed in more details in the following section.

The friction landscape was obtained by scaling the GTOPO30 Digital Elevation Model (DEM) [150] for our area of interest. Higher friction values correspond to lower values of the scaling function  $\sigma$  of the diffusion process and vice versa. The scaling function  $\sigma$  is bound from below by zero, but also from above by a constant  $\sigma_{\max} > 0$ . The mobility model is calibrated using mobility data of contemporary hunter-gatherers in Central Africa.

We choose the radius for foraging and short range interactions  $r_1$  to be 20km based on estimations in [144]. The carrying capacity of the landscape was calculated for our choice of  $r_1$  following the approaches in [111, 144]. Based on population statistics of present day hunter-gatherers [111] we chose the thresholds for fission and fusion to be  $h_{\text{fis}} = 60$  and  $h_{\text{fus}} = 18$ . The rates for population growth and decline are chosen as  $\rho_g = \rho_d = 0.001$ , which corresponds to 0.1% per year and is within range of estimates for prehistoric population growth rates for hunter-gatherer tribes [151].

For the rate constants of the cultural evolution we will explore different scenarios in the next section and choose the rate constants to be independent of the feature index  $i$ . We measure the time in years and set the default time-step size for our model to be 1 month.

### 3.2.2 Computational Analysis

To analyze our model outcomes we first have to think about which macroscopic and mesoscopic quantities are of interest. On the mesoscopic scale we can identify spatio-temporal agent clusters with a more densely connected interaction network and investigate how these spatial clusters can be related to the cultural similarity of the agents. Therefore we will consider two clustering approaches. The first one is taking only the positions of the agents into account resulting in a clustering that only depends on the agent mobility. The second approach defines the clusters purely based on the cultural status of the agents and only after the clusters are assigned to the agents we can make the connection to spatial areas and a comparison of the clustering approaches.

With regards to the cultural dynamics of our model the arrival times of a specific cultural status is at least in this general setting not very meaningful. We are in our setting more interested in measuring the cultural similarity and cultural differences of the agents. One way to quantify these on the macroscopic scale is the utilization of a diversity index. Our measure of choice will be a variant of Simpson's diversity index that is defined by

$$I(Y(t)) = 1 - \frac{\sum_{i=1}^{R(Y(t))} n_i(Y(t))(n_i(Y(t)) - 1)}{n(Y(t))(n(Y(t)) - 1)}, \quad (3.11)$$

where  $R(Y(t))$  is the number of different status values of the system state  $Y(t)$  and  $n_i(Y(t))$  the number of active agents with the same status for each of the observed status vectors and  $n(Y(t))$  is the total number of active agents at time  $t$ . By construction we have  $I(Y(t)) = 1$  if all agents have a different cultural status and  $I(Y(t)) = 0$  if the status of all agents is the same.

In the case of progressive cultural dynamics the value of the cultural status of the agents has an interpretation as a level of complexity and the average cultural complexity of the agents is a meaningful macroscopic variable to study. We will consider the development of the average cultural complexity over time for different scenarios and we can on the one hand observe the impact of different parameter settings on the results and on the other hand relate the changes in cultural complexity to changes of the suitability landscape.

While all three observables and their changes in relation to different parameter settings can also be interpreted in an anthropological sense [17], we will be mostly concerned with the dynamical characteristics of the different scenarios. We will not construct a reduced model as we did for the first application but we will briefly discuss the challenges of model reduction with regards to the more complex ABM dynamics, especially the consequences of a changing landscape and temporal clusters.

#### Cultural diversity

In Figure 3.17 we illustrate the development of our chosen diversity index over time for both types of cultural dynamics with similar parameter settings (the only difference is

the frequency of first-order events, which is in the non-progressive case chosen to be an order higher). We can observe that the diversity is higher in the progressive case despite the lower frequency of first order events and that both curves are highly correlated. This is on the one hand due to the nature of the transmission dynamics that lead in both cases to an increase of similarity between the agents that are connected in the interaction network. While in the non-progressive case within spatial agent clusters the transmission of more frequent status values is more likely in the progressive case the transmission of higher status values is favored. Thus, while in both cases there is a drift towards a consensus within agent clusters in the progressive case there is more often a change of which status value is representing this consensus of a cluster. This leads to the overall higher diversity in the progressive case.

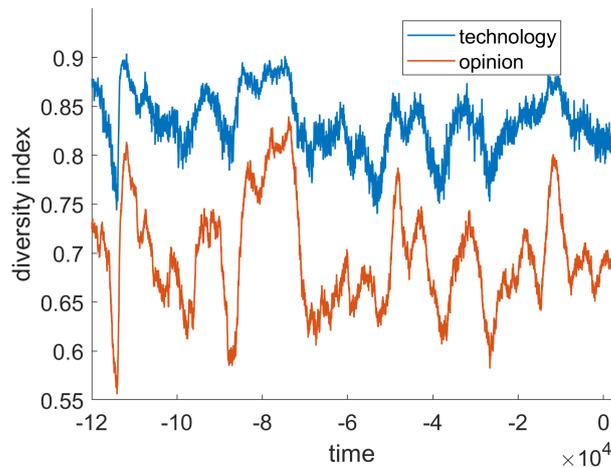


Figure 3.17: Cultural diversity over time according to the index defined in Equation (3.11) for both progressive (technology) and non-progressive (opinion) cultural dynamics. The results are averaged over 20 simulations for each of the two scenarios.

On the other hand in both cases we have a strong dependence of the transmission dynamics on the distribution of the agents in space. Substantial changes in the suitability landscape can lead to vastly different interaction networks. This is also reflected in the diversity index, the strong changes in overall diversity are in both types of cultural dynamics highly correlated and can be linked to big changes in the suitability landscape. Greater increases of the diversity index can for example be linked to more fragmented landscapes that feature a higher number of clusters or an overall decreased carrying capacity leading to lower numbers of active agents. Conversely a change to a landscape that leads to agents being concentrated in only a few large clusters is connected to stronger decreases of the diversity index. While on the one hand we can explain strong changes in diversity with substantial landscape changes we can on the other hand also

use the strong changes in diversity as an indicator to identify substantial changes in the landscape.

### Parameter sensitivity

When varying the parameters that specify the mobility process, i.e., the parameters specifying the scaling function  $\sigma$  of the diffusion, as well as parameters for weighting the three different influences small changes of the parameters lead only to minor differences in the mobility patterns. Despite also having slight differences in the distribution of the agents we consider mostly statistics about travel distances to quantify the differences of the scenarios. We measure the average travel distance per default time step, the average number of residential movements per year, i.e., movements with a travel distance of at least 2 km as well as the average total distance travelled per simulation year. The residential movements as well as the travel distances are also the mobility characteristics that we can compare to the real-world data of contemporary hunter gatherer mobility to assess whether the mobility process of our model is plausible. As we have no reliable data on prehistoric hunter gatherer mobility a validation in the classical sense is not possible. The small variation of the mobility parameter values does not change the characteristics of the cultural dynamics significantly, i.e., the overall diversity and accumulation of complexity is not affected by small parameter changes. Thus, we choose one fixed plausible setting for the mobility parameters when we investigate the sensitivity of the parameters for demographics and cultural dynamics.

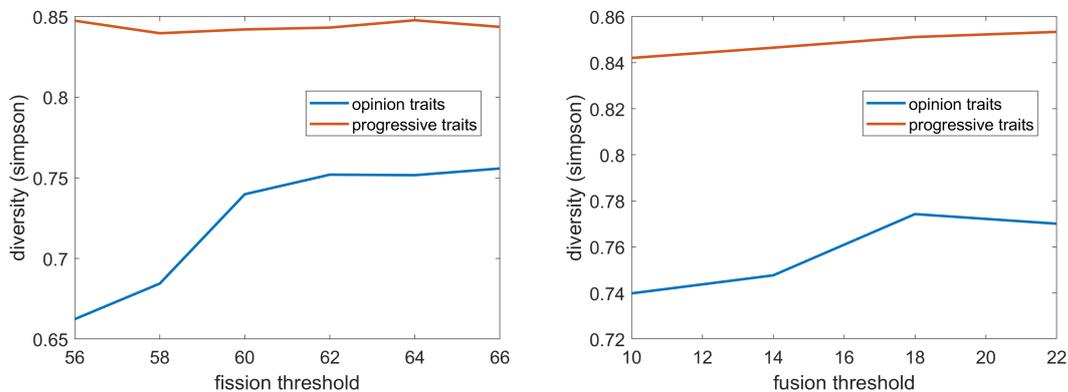


Figure 3.18: Relationship between fission and fusion thresholds and Simpson's diversity for a high frequency of first order events. The curves for the non-progressive dynamics are labeled as opinion traits in the plots. The results for each parameter setting are averaged over five shorter simulations.

For the parameters of the demographic dynamics, i.e., the fission and fusion thresholds we consider four general scenarios. We distinguish between progressive and non-progressive

cultural dynamics with both a high and a low frequency of first order events. The relations between the threshold parameters and the cultural diversity are rather similar in all four scenarios and visualized for the case of a high frequency of first order events in Figure 3.18. It is not surprising that higher values of  $h_{\text{fus}}$  and  $h_{\text{fis}}$  both lead to a lower number of agents and a higher number of camp members per agent. Higher numbers of agents are in general correlated with a decrease in diversity which can be explained by a higher connectivity of the interaction network and thus a faster convergence to a consensus within the agent clusters. Thus both thresholds are positively correlated with cultural diversity in the model and as we can see in Figure 3.18 there is a saturation effect in the non-progressive case which serves as an additional justification for the parameter choices made. As higher agent numbers are leading to a higher frequency of innovation events in the progressive case there is in general a lower average complexity for higher values of  $h_{\text{fus}}$  and  $h_{\text{fis}}$ .

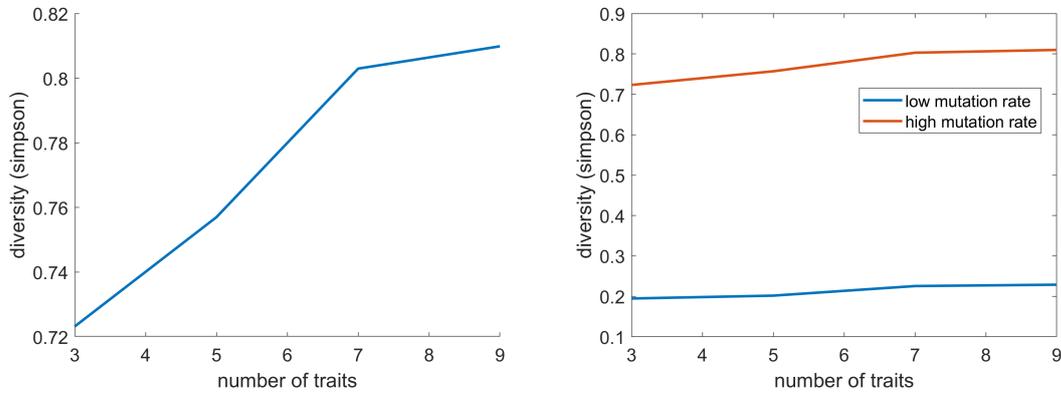


Figure 3.19: Relationship between the number of possible trait values for non-progressive dynamics and diversity. The left plot is showing the results for the case of a high frequency of first-order events.

The parameters for the cultural dynamics are the rate constants for each type of adoption event and in the non-progressive case also the number of possible trait values. Increasing or decreasing all rate constants by the same factor only changes the relation between the cultural dynamics and the mobility dynamics. As long as the changes are not too drastic the diversity is not affected much, while the accumulation of complexity increases with overall faster cultural dynamics. More interesting are the relations between the different rate constants, especially in the progressive case. In the non-progressive case an increased frequency of first order events in relation to the frequency of transmission events also increases the diversity and vice versa. The diversity is increasing with the number of possible trait values  $m$  but only for low values of  $m$  (compare Figure 3.19). Once there are sufficiently many available possible trait values a further expansion of the status space does not lead to significantly more diversity. So while we could for increased

realism also expand the status space in the non-progressive case this would not lead to more diversity or additional cultural clusters, as already for our example simulation with  $m = 10$  possible trait values there is only a small number of trait value combinations associated with distinct clusters out of 1000 possible unique status vectors.

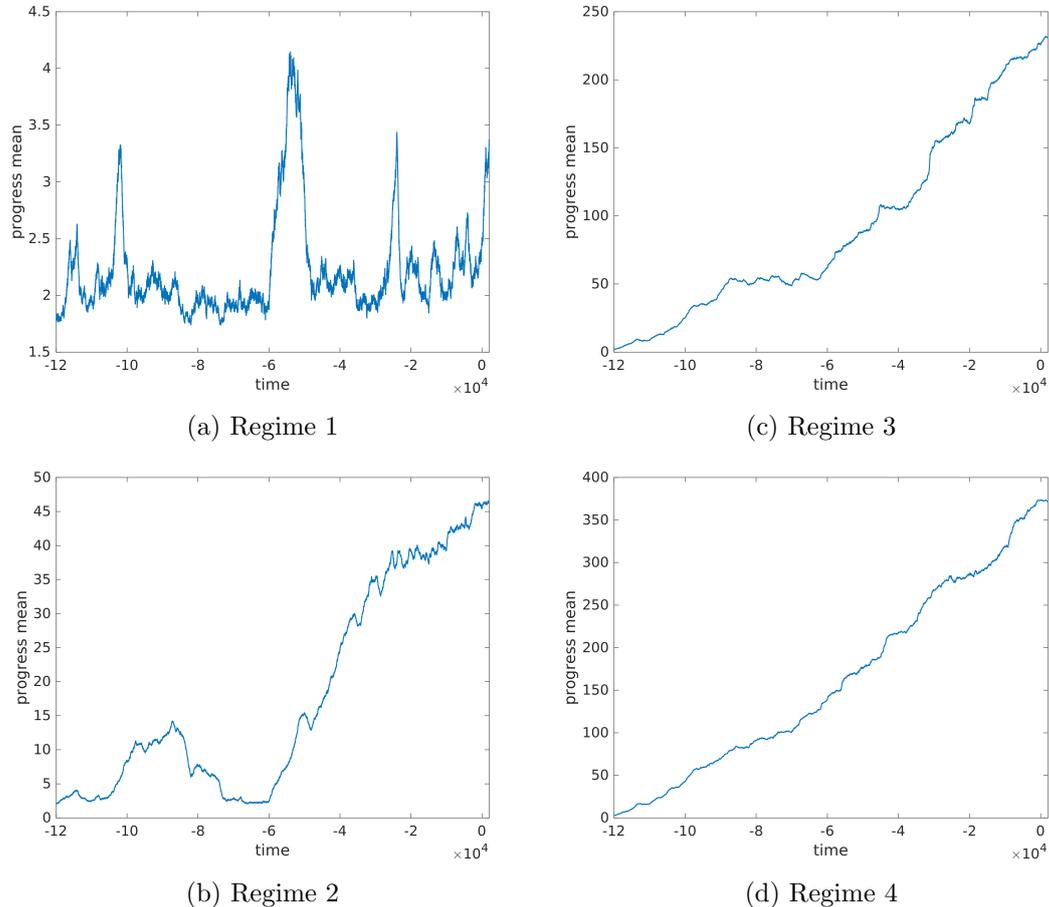


Figure 3.20: Average cultural complexity over time for four different choices of interaction rates.

In the progressive case we have two types of first order adoption events to consider, gain as well as loss of information. At first we consider the rates for gain and loss of information to be equal. In this case we have a steady increase in the average cultural complexity and can observe a diversity pattern similar as in Figure 3.17. An increased rate for information gain  $\gamma$ , while fixing the loss rate  $\lambda$  and transmission rates  $\varphi_1, \varphi_2$ , leads to an increase in diversity and a faster accumulation of cultural complexity. If we fix the gain rate  $\gamma$  and the transmission rates  $\varphi_1, \varphi_2$  and vary the loss rate we can differentiate between different parameter regimes. Small increases of  $\lambda$  do not affect the overall characteristics of the dynamics. Larger increases of  $\lambda$ , however, lead to

time frames where the average trait values representing cultural complexity are remaining constant or even decreasing and not only increasing. These phases are related to changes in the suitability landscape that lead to reduced connectivity in the interaction network. Very large increases of  $\lambda$  even lead to cultural complexity only rising for short spikes or even not at all. The intermediate parameter regime where the accumulation of cultural complexity depends stronger on the landscape changes begins when  $\lambda$  is about two orders of magnitude lower than  $\gamma$ . Considering this ratio between  $\gamma$  and  $\lambda$  as fixed then we can find similar parameter regimes for variations of the interaction rates  $\varphi_1$  and  $\varphi_2$  (see Figure 3.20). The characteristics visualized in Figure 3.20 (b) can be observed when the interaction rates are chosen such that the overall frequency of loss events is similar to the combined frequency of gain and transmission events. In this case only for suitability landscapes that lead to interaction networks with higher connectivity the cultural complexity is increasing. In the anthropological sense this can be interpreted as a higher connectivity and more frequent exchange between different hunter-gatherer camps leading to a prevention of information loss. The parameter settings with very high information loss rates are requiring more computational effort and because of the high frequency of events the average step size is significantly smaller than the default step size. In this situation a realization of the simulations utilizing a tau-leaping method or a hybrid method instead of the purely event-based algorithm could lead to a gain in computational effort.

More details on the parameter sensitivity, the case study and the anthropological interpretation of the model can be found in [17].

### Identifying spatio-temporal clusters

The agents of our model distribute according to the attractive areas of the suitability landscape and form spatial clusters. By the definition of the cultural evolution dynamics also a clustering in the status space emerges that has strong ties to the spatial clustering. For the spatial cluster detection we can apply the hierarchical density based clustering algorithm HDBSCAN for a time frame in which the suitability landscape does not change, just like we did for the construction of the PDMM approximation of the ABM for innovation spreading. However, we need to deal with several challenges when we want to identify clusters based on the status of the agents instead.

The key to any clustering method is to define measures for closeness, connectivity or density. In our case we want to define a clustering of the landscape that depends on densely connected agents within the area being close in status. Through the nature of our cultural status dynamics we have a lot of noise through first order events and depending on how we define closeness in status even single mutation or loss events can already be considered to be a strong perturbation. Thus, closeness in status has to be considered in a way that the perturbations through first order events are filtered out in some sense, while still allowing for a strong enough distinction between agents such that we can still identify more than just one large cluster. For that we consider connectivity

to be defined over a fixed time period that is long enough such that a few noisy snapshots in the data do not hold a strong influence on the outcome of the connectivity calculation.

We define for each system state  $Y(t)$  an adjacency matrix  $\hat{A}(Y(t))$  with entries

$$\hat{A}_{\alpha\beta}(Y(t)) := \begin{cases} 1, & \text{if } \|s_\alpha(t) - s_\beta(t)\| \leq h_a \\ 0 & \text{else} \end{cases}$$

for a suitable threshold  $h_a \geq 0$ , that defines closeness in status between two active agents  $\alpha$  and  $\beta$ . Entries referring to inactive agents are defined as 0. In the non-progressive case, we set  $h_a = 0$  and thus consider agents only close in status if their status is exactly the same.

Next we consider a set of consecutive system state snapshots  $Y(t_1), \dots, Y(t_M)$  and define the time averaged adjacency matrix  $B([t_1, t_M])$  with entries

$$\hat{B}_{\alpha\beta}([t_1, t_M]) := \frac{1}{M} \sum_{k=1}^M \hat{A}_{\alpha\beta}(Y(t_k))$$

The entries of  $\hat{B}([t_1, t_M])$  can be interpreted as the percentage of the time interval  $[t_1, t_M]$  that two agents can be considered close in status with respect to the definition of closeness via  $\hat{A}$ .

We define the time averaged connectivity matrix  $C([t_1, t_M])$  with entries

$$C_{\alpha\beta}([t_1, t_M]) := \begin{cases} 1, & \text{if } \hat{B}_{\alpha\beta}([t_1, t_M]) > h_c \\ 0 & \text{else} \end{cases}$$

with connectivity threshold  $h_c \in [0, 1]$ . The connected components of  $C([t_1, t_M])$  correspond to cultural clusters of agents that are close in status for a significant amount of the time interval  $[t_1, t_M]$  with the threshold  $h_c$  specifying the fraction of the time period. Each agent  $\alpha$  can then be assigned a label  $\ell_\alpha(t) \in \mathbb{N}$  corresponding to the connected component that it belongs to at time  $t$ . We define a minimum cluster size  $h_m$  and assign for agents belonging to cultural clusters with less than  $h_m$  members the label 0.

The lower the connectivity threshold  $h_c$  is, the more noise in the status variable is allowed within the same cultural cluster. For larger values of the threshold  $h_m$  the cultural clusters need a larger size in order to be assigned a distinct label. In our illustrative example we have chosen  $h_c = 0.5$  which means that more noise is allowed, but we still can interpret the results such that within the cultural clusters the agents are close in status for the majority of the snapshot sequence. This interpretation would be lost for the choice  $h_c < 0.5$  that would allow for more noise within the cultural clusters. In general, lower values of  $h_c$  allow for more connections between agents in  $C([t_1, t_M])$ . For very large values of  $h_c$ , in the example simulations we observe a lower number of cultural

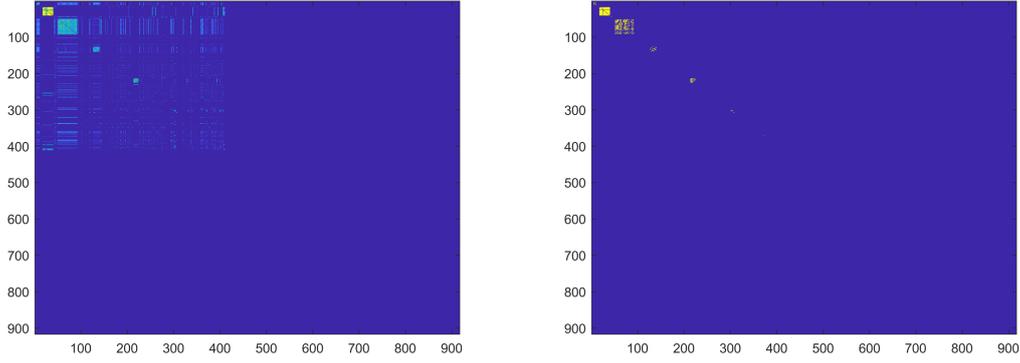


Figure 3.21: Illustration of the matrices  $B([t_1, t_M])$  (left) and  $C([t_1, t_M])$  (right) for a time period  $[t_1, t_M]$  with static suitability landscape. Only entries associated with agents that have been active in the time period have non-zero entries. The left matrix has entries between 0 (dark blue) and 1 (yellow), the right matrix has entries that are either 0 or 1.

clusters. We have chosen  $h_m = 5$  for our example, so we also allow for the identification of cultural clusters of smaller size. To not identify single agents as a distinct cultural cluster one should set at least set  $h_m > 1$ .

Next, we will introduce a way to determine how spatial areas can be associated with the identified cultural clusters. Through the trajectory data in the time frame  $[t_1, t_M]$  with  $M$  system state snapshots we can translate the cultural clusters to areas of the suitability landscape. We define a grid  $\mathcal{G}$  with a suitable resolution that is covering  $\mathbb{X}$ . For a grid cell  $G \subset \mathbb{X}$  we define the total number of visits of agents with label  $j > 0$  as

$$v_G(j) := \sum_{\alpha} \sum_{k=1}^M \chi_j(\ell_{\alpha}) \chi_G(x_{\alpha}(T_k)).$$

Using this quantity, we cluster all grid cells, such that we assign to a grid cell  $G$  the label of its most frequently visited cultural cluster

$$\ell_G := \begin{cases} \operatorname{argmax}_j \{v_G(j)\}, & \text{if } \max_j \{v_G(j)\} > h_l \\ 0 & \text{else} \end{cases}$$

with a threshold  $h_l \geq 0$  specifying the number of agents' visits for becoming a part of a cultural cluster. If a grid cell was not visited often enough by any agent that belongs to a specific cultural cluster, it gets assigned the value 0. An assignment to 0 can be interpreted in multiple ways: one case is that the grid cell belongs to a transition region that is not densely populated with agents. Alternatively, there could be multiple agents in an area, but their interaction network is not densely connected enough to establish a similarity in status.

**Remark 3.** Due to the stochastic nature of our ABM, the outcome of the cultural clustering depends strongly on the realization. Thus, although the spatial distribution of agents according to the mobility process is similar throughout multiple realizations, the emergence of cultural clusters can be vastly different for each trajectory. Therefore, in order to analyze the cultural clusters more thoroughly, it would be reasonable to consider statistics of the clustering results over multiple simulation runs.

Applying both methods, HDBSCAN for the mobility data and the in this section introduced method for identifying clusters based on status similarity, we are able to identify the mobility clusters and the cultural clusters of our model. For that we choose time frames that match the time periods in which the suitability landscape is static. Both types of clusters are positioned at the suitable areas of the landscape of that time period and more fragmented suitability landscapes lead in general to a higher number of (mobility) clusters. In the illustrations of the spatio-temporal clustering (Figures 3.22 and 3.23) we can observe singular cultural clusters including multiple mobility clusters as well as singular mobility clusters containing multiple cultural clusters. Two mobility clusters though spatially separated can still be dynamically close in the sense that transitions between the mobility clusters occur frequent enough, such that the agents of both mobility clusters still possess a similar cultural status. On the other hand a large mobility cluster can be structured such that the interaction network is modular and thus multiple cultural clusters can emerge due to the timescale separation.

We also can observe the dynamics of the different clusters over time, as we can connect the different clusterings via the trajectories of the agents as is illustrated in the alluvial diagrams of the Figures 3.22 and 3.23. In general, we can distinguish between three different types of cluster dynamics that are typically considered for the analysis of dynamic communities [152]:

1. merging and splitting of clusters
2. emergence and disappearance of clusters
3. expansion and contraction of cluster areas.

These dynamical patterns are not necessarily mutually exclusive, e.g., a cluster can split into multiple separate clusters and in addition there are also changes to the shapes of the clusters through expansion and contraction of cluster areas. We can observe all of the above patterns in the illustrations of Figures 3.22 and 3.23. In Figure 3.23 we have clusters 4 and 5 of the first time period merging into cluster 3 of the second time period and then splitting again into clusters 4 and 5 of the third time period. In the same figure cluster 3 of the first time period is an example for disappearance of clusters and cluster 3 of the third time period an example for an emerging cluster. Considering our two types of cultural dynamics, in general we can observe a higher number of distinct cultural clusters in the case of progressive dynamics. In the non-progressive case we have more often the situation that the splitting of a mobility cluster does not induce the splitting of a cultural cluster (see Figure 3.22). This is because the non-progressive

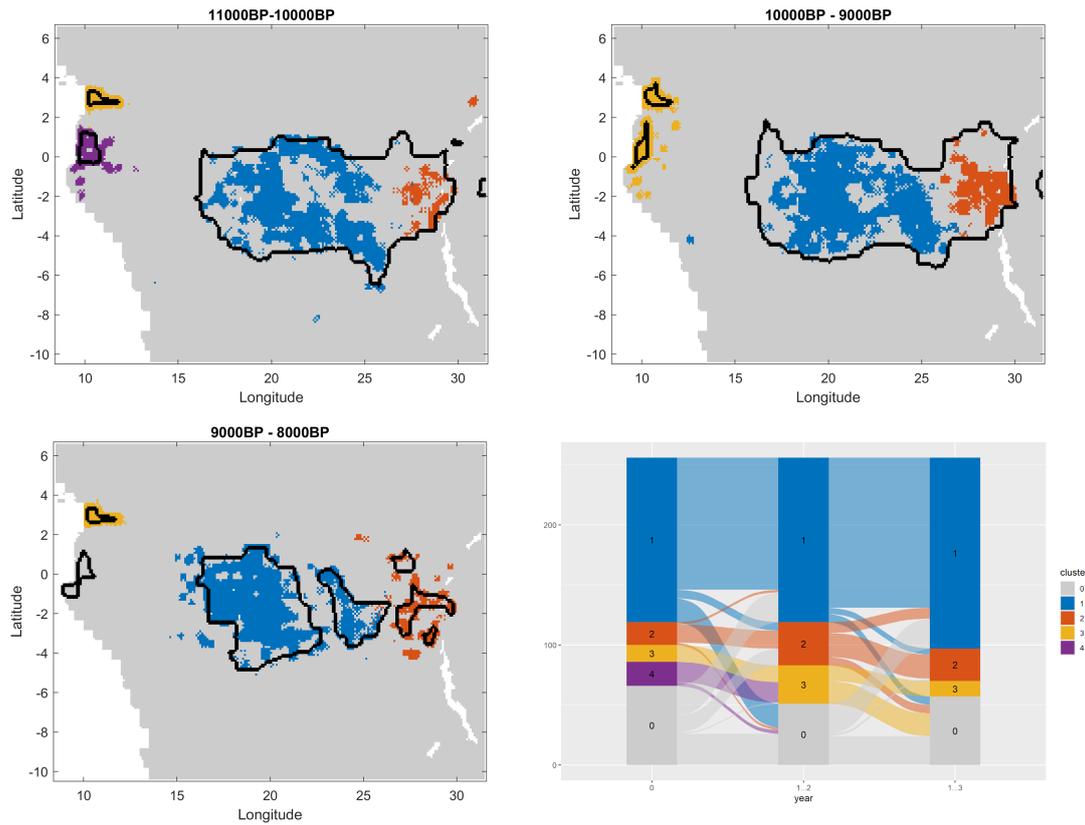


Figure 3.22: Comparison between clusters based on non-progressive cultural status (colored areas) and clusters based only on mobility trajectory data (marked by black borders) for three consecutive suitability landscapes centered around 10000 BP. The relations between the different cultural clusters are illustrated in an alluvial diagram.

dynamics tend to maintain a specific value for a status consensus among agents, while in the progressive case higher trait values are preferred and thus the spatial separation leads more quickly to a difference in cultural status.

The temporal relations between the clusters are on the one hand also very interesting for an anthropological interpretation of the model [17] and on the other hand also important for the possible construction of a reduced model of the ABM on the mesoscale.

### Model reduction

In the case of an autonomous ABM we can identify the metastable sets of the state space and apply the model reduction framework of Chapter 2. For the hunter-gatherer model we have an additional dependence of the generator on time and we have observed in the mobility clustering results that an area of the state space that would be metastable with

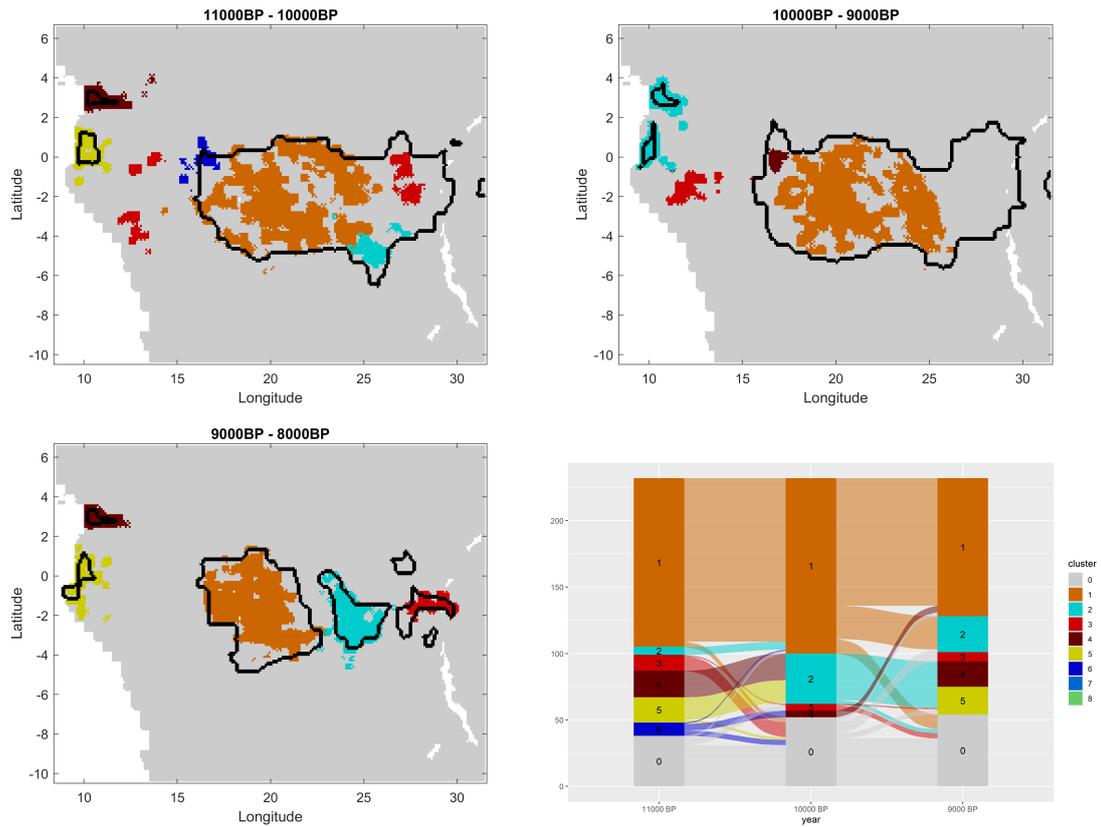


Figure 3.23: Comparison between clusters based on progressive cultural status (colored areas) and clusters based only on mobility trajectory data (marked by black borders) for three consecutive suitability landscapes centered around 10000 BP. The relations between the different cultural clusters are illustrated in an alluvial diagram.

respect to the mobility process at time  $t_1$  is not necessarily metastable at time  $t_2$  as well. There are several approaches to this problem, one of them would be to expand the state space with the time variable such that we can simply apply the concept of metastable sets on this expanded state space [153]. Another approach is the generalization of metastable sets to a coherent family of sets [154]. The notion of a coherent pair consisting of two sets  $A_1$  and  $A_2$  is a good starting point for understanding how the notion of coherence is connected to metastability and almost invariant sets. We call  $(A_1, A_2)$  a coherent pair if on the one hand the probability  $\mathbb{P}(\mathbf{X}(t_2) \in A_2 \mid \mathbf{X}(t_1) \in A_1) \approx 1$  and on the other hand  $\mathbb{P}(\mathbf{X}(t_1) \in A_1 \mid \mathbf{X}(t_2) \in A_2) \approx 1$ . Figuratively speaking this means that a process starting in  $A_1$  is very likely to end up in  $A_2$  and a process that ends up in  $A_2$  has been in  $A_1$  before with a very high probability. This forward-backward relationship can be utilized to define a process by the composition of the forward transfer operator and its adjoint, i.e., the backward transfer operator, and with respect to this forward-backward

process [154] a coherent pair would be a metastable set with respect to the time scale defined by the two time points  $t_1$  and  $t_2$ . A coherent family is then the expansion of coherent pairs to a time-indexed family of sets with any two arbitrary sets of the coherent family fulfilling the property of a coherent pair. Depending on the process it makes sense to only consider coherence on a finite time scale that is significantly faster than the mixing time of the process. Otherwise, the case may arise that only the complete state space can be considered a coherent set. The identification of coherent sets can then also be done via a spectral approach by analyzing either the transfer operator of the process on the expanded state space or the transfer operator of the forward-backward process [153, 154]. A discretized model based on a projection utilizing the coherent families of the process could then be constructed and several discretization approaches for the transfer operator are presented in [153].

For our model it would make sense to utilize the stationary time periods during which the generator of the process is not depending on the time variable. For these time periods we can construct a stochastic metapopulation model according to the framework of Chapter 2 as our more complicated status dynamics are still only composed of first and second order adoptions. For the choice of the subpopulations we can consider on the one hand the mobility clusters or on the other hand the cultural clusters. Both approaches could be justified, considering that both clusterings are closely related and both describe a metastable behavior with respect to the process dynamics. A comparison of the different approaches would then also be interesting to investigate.

A first challenge is how to deal with the notation of the multidimensional cultural status and the additional continuous population variable that is not discretized. While we could denote the system state of the reduced model in a multi-dimensional matrix  $N$  in the non-progressive case, we would have to further adapt the notation for the progressive case. Accounting for all possible status values would lead to  $N$  being a countable matrix and the object encoding the system state would then be actually of higher dimension than the original ABM system state. Also in the non-progressive case there are in total more possible status values than agents. Thus, the most reasonable choice might be to simply denote the system state of the reduced model in the same fashion as the ABM system state  $Y$  but with the set of subpopulation labels replacing  $\mathbb{X}$  in the system state space. The construction of the projected adoption rate functions has then to be adapted accordingly. Considering the transitions between the subpopulations we would have a change in subpopulation assignment for a chosen agent when a transition event happens.

The demographic dynamics are deterministic by construction but due to the dependence on the stochastic mobility process the population variable exhibits stochastic characteristics. One possibility to deal with the demographics would be to let number of members of all camps of a subpopulation grow or decline simultaneously depending on whether the total number of camp members in a subpopulation is lower or higher than the total carrying capacity of the area associated with the subpopulation. Fission and fusion of the camps could then be done in the same way as in the ABM while considering every

camp of the subpopulation as close enough in case of a fusion event. The stochasticity of the camp population variable would, however, be lost as the total number of camp members would quickly converge to the total carrying capacity of the subpopulation and then no longer change significantly.

A further reduction to a PDMM would not reflect the more stochastic nature of the cultural dynamics well. In this case it would make more sense to consider a diffusion approximation for the internal dynamics of the subpopulations as is done in the context of the chemical Langevin equation. An example for the case of opinion dynamics on a complete graph can be found in [155]. The agent numbers of our model however do not require the additional approximation for effort reduction and we would be happy to just stick with an SMM.

Thus, we have an idea how we could construct a reduced model for the stationary time periods. The question that remains is how to connect the different metapopulation models at the time points where the suitability landscape changes. Considering that we define the subpopulations via clustering it makes sense to use the temporal relations between the clusters to define the relations between the subpopulations of the different models. As we track for each agent the subpopulation assignment we can simply reassign the subpopulation of all agents according to the statistics that we gathered from the trajectory data. An assignment to the label 0 could then for convenience be equivalent to the agent becoming inactive. This way we already have covered the cases of merging/splitting and disappearance of subpopulations. The change of the cluster area through expansion/contraction is reflected in the change of the total carrying capacity. What is still unclear is how to deal with the emergence of a subpopulation. We could assign to an emerging subpopulation a sufficient number of previously inactive agents such that the total number of camp members reflects the total carrying capacity. For the initialization of the status variables of the agents a random initial state according to a uniform distribution would make the most sense in the non-progressive scenario. In the progressive case, however, the initial trait values should be at most the minimum trait values present in all other status vectors. Otherwise, a transition from an agent of the newly emerged subpopulation could directly influence the cultural complexity of formerly established clusters.

This is, however, only a rough sketch of how to approach the challenge of model reduction for this more complex ABM system. To actually build a reduced model and further investigate these and alternative assumptions could be the topic of future research.

### 3.3 Epidemic Spreading

Of all types of spreading processes, the spreading of infectious diseases has probably the longest history, with early models dating back to the 18th century [156], and certainly received the most attention in the last few years due to the COVID-19 pandemic. Suddenly, there was a strong demand for accurate models for the spreading of the SARS2-CoV to assist in the decision-making process for finding a response strategy. But also the amount of available data, e.g., on case statistics, has been unprecedented, providing an opportunity for the development and evaluation of new models. The variety of modeling approaches and methods that have been applied to the problem is as vast as the state of the art allows. Most prominent are still macroscale approaches such as compartmental ODE models, ranging from classical SIR models [157] to many added compartments, e.g., for quarantine or hospitalization states [158], as the parameters for these models can be estimated well from the available data on case statistics and thus can be quickly adapted to new situations. However, also many mesoscale approaches like network [159] or metapopulation models [160] have been applied as well as agent-based models on the microscale [161]. While these approaches usually require more assumptions, additional data and can be more difficult to parameterize they allow for the discussion of more specific research questions that are directed at the core dynamics of infection spreading, e.g., about the effectiveness of different kinds of non-pharmaceutical intervention measures.

In this section, we will briefly discuss the basic assumptions of standard modeling approaches for epidemic spreading within present day societies for different spatial resolutions, especially those concerning the mobility of the population members. Afterwards, we take a closer look at a piecewise-deterministic metapopulation modeling approach to epidemic spreading on the mesoscale with adaptive rate constants, to e.g., model the implementation of non-pharmaceutical interventions. At the end of the section, we will briefly discuss under which conditions we could apply the model reduction framework from Chapter 2 to an ABM for epidemic spreading.

### 3.3.1 General Model Assumptions for Epidemic Spreading

The spreading of infectious diseases within human populations is a very complex process that depends on many interactions on various spatial and temporal scales. No useful model will ever be able to capture all the details of the real-world process at once as many subprocesses are already on its own challenging modeling problems. To only name a few, there is the interaction of the virus with the host [162], the transmission of the virus between two individuals [163], the behavior and mobility of individuals within the society [164]. The larger the scale of the model is, the more we need to simplify the details of the dynamics to be able to actually simulate or analyze the model dynamics. Thus, for the epidemic spreading on a population scale, we usually make very strong simplifications of the microscopic interactions between the virus and the individuals. Instead of modeling the details of the virus-host interactions and transmission dynamics, we assume that the course of the disease of an individual can be modeled by stochastic jumps between various expressions of an epidemiological status.

#### Compartmental ODE models

The common basis of most epidemic models distinguishes between three different status expressions: An individual can be either susceptible to the disease ( $S$ ), infected with the disease and able to transmit it ( $I$ ) or removed from the infection dynamics ( $R$ ), which in general can also be interpreted as recovered from the disease. The transition from status  $I$  to status  $R$  is usually modeled by a first order adoption of the new status with a constant rate, strongly simplifying the actual course of the disease, and the switch from  $S$  to  $I$  modeled by a second order adoption due to pairwise interactions with an infection rate constant, which is a strong simplification of the transmission dynamics. On the coarsest spatial scale, we simply assume that the population is *well-mixed*, i.e., all individuals have the same probability to interact with each other. We do not consider individual infection and removal rates but rather general rates that reflect average rates that can be derived from data, and thus we can write down the state of the system similar as we did for the metapopulation models in Section 2.2, i.e., we divide the population  $N$  in compartments  $N_S, N_I, N_R$  according to the infection status. The resulting model is known in the literature as *SIR* model for epidemic spreading and can either be formulated as a jump process or deterministically. The formulation of the jump process for the model can be written as

$$\begin{aligned} N_S(t) &= N_S(0) - \mathcal{P}_{SI} \left( \int_0^t \gamma_{SI} N_S(s) N_I(s) ds \right) \\ N_I(t) &= N_I(0) + \mathcal{P}_{SI} \left( \int_0^t \gamma_{SI} N_S(s) N_I(s) ds \right) - \mathcal{P}_{IR} \left( \int_0^t \gamma_{IR} N_I(s) ds \right) \\ N_R(t) &= N_R(0) + \mathcal{P}_{IR} \left( \int_0^t \gamma_{IR} N_I(s) ds \right) \end{aligned}$$

with  $\gamma_{SI}$  being the infection rate constant,  $\gamma_{IR}$  the removal rate constant and  $\mathcal{P}_{SI}, \mathcal{P}_{IR}$  independent unit rate Poisson processes.

As we are modeling on the macroscale in general we take into account a sufficiently large population and thus we reasonably can consider the deterministic population limit for the dynamics instead. The corresponding ODE system can be written as

$$\begin{aligned}\frac{d}{dt}\mathbf{N}_S(t) &= -\gamma_{SI}\mathbf{N}_S(t)\mathbf{N}_I(t) \\ \frac{d}{dt}\mathbf{N}_I(t) &= \gamma_{SI}\mathbf{N}_S(t)\mathbf{N}_I(t) - \gamma_{IR}\mathbf{N}_I(t) \\ \frac{d}{dt}\mathbf{N}_R(t) &= \gamma_{IR}\mathbf{N}_I(t)\end{aligned}$$

with the same rate constants for infection and removal as in the stochastic model. This formulation as a *compartmental ODE model* is the basic model that is widely used for epidemic spreading on the macroscale and has the advantage that the model parameters can be derived from observed case statistics.

While there are many extensions with additional compartments to more accurately model the spread of a particular disease in a specific society, the general dynamics do not change significantly and the strong simplification of the mobility remains with all its consequences for applicability and interpretability. While the well-mixed assumption might hold on the level of cities, and even there some spatial patterns can be observed [165], it certainly is too strong for modeling the spread between countries on a global scale. But also on the scale where the mobility can reasonably be simplified in this way, the simplification of the societal dynamics is also an issue.

For example, the infection rate of compartmental ODE models incorporates many different factors of the original process. It reflects the probability that a contact between two individuals leads to a transmission as well as the average number of contacts an individual has over time due to the current behavior and mobility of members of the population and is thus not an intrinsic property of a virus or other vector but also depending on the society in which it spreads, e.g., for cities or countries with different population densities and demographics the estimated infection rates for the spreading of the same disease can be different [166]. A variation of the infection rate constant can then also have many different interpretations, e.g., a reduced infection rate can be interpreted as a reduced transmissibility of the virus due to wearing face-covering masks or a reduced number of contacts either due to restricting policies or a generally more cautious behavior.

So while a compartmental ODE model might be suitable for short term predictions of the case statistics of a chosen population, for discussion of more detailed questions other approaches are better equipped. For the spreading on a larger scale, where the well-mixed assumption does not hold, a metapopulation model can be used to assess the risk of spreading between different cities or countries. For the detailed analysis of specific measures to prevent the spreading of the disease within the population, an ABM on the microscale might be the model of choice.

### 3.3.2 PDMM for Epidemic Spreading

The term metapopulation model was first introduced in the context of ecology [167] to model the environmental heterogeneity of spatially separated subpopulations. Since then, metapopulation models have been widely used for different types of population dynamics characterized by a timescale separation between the dynamics within and between the different subpopulations [67]. Besides the spatial separation, different social groups (e.g. based on age) have also been considered for the division of the population in epidemiological models [168]. Purely deterministic metapopulation models, in which all flows between compartments within subpopulations as well as the interaction between subpopulations are defined by an ODE system, as well as purely stochastic approaches, both have their advantages, either in efficiency or in realism. A piecewise-deterministic approach to metapopulation models can be advantageous compared to both of the pure approaches. Especially in the case of multiscale dynamics, the PDMM has the advantage of being more efficient than a purely stochastic model, especially for large populations, as well as more realistic than the purely deterministic model for the slower processes that are modeled with stochastic jumps. For modeling the epidemic spreading between spatially separated subpopulations we can assume that there is a significant time scale difference between the dynamics of travelling between subpopulations, which is necessary to enable the spreading of a disease between members of different subpopulations, and the interactions of individuals within a subpopulation. Thus, a piecewise deterministic metapopulation model is a good choice for modeling epidemic spreading on a spatial scale where the well-mixed assumption is no longer reasonable.

We will now apply the PDMM approach to model the spreading of COVID-19 between two spatially separated subpopulations. This conceptual model has already been presented in [16] and it will be presented here in a similar manner including the figures and some text passages from the original publication.

#### PDMM for spreading of COVID-19

In the case of COVID-19, metapopulation models provide a good approximation of the original epidemic dynamics since the metastability assumption can be observed in mobility data [169], e.g., in rare spatial transitions between different cities or countries caused by mobility restrictions. Moreover, sufficiently large population sizes are realistic to justify the assumption of piecewise-deterministic dynamics and to allow for a significant reduction in model complexity compared to a purely stochastic metapopulation model or a microscale ABM. Although we choose to construct a conceptual model, the PDMM can be easily calibrated to real-world data, as we have good estimates for rates at the population scale.

For simplicity, we consider two subpopulations that have frequent local interactions

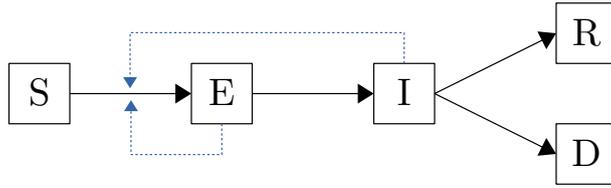


Figure 3.24: Visualization of the SEIRD model. Black arrows stand for possible status transitions, blue arrows indicate an impact by interaction.

and rare transitions between them. The model was calibrated to parameters estimated in studies from the beginning of the pandemic. Note that the model is not applied to analyze a specific real-world dataset, but representative results are used to show how the model can be applied to possible real-world scenarios. In addition, we will analyze the effectiveness of various containment measures taken within subpopulations and global measures that affect the spread between subpopulations. However, the main goal of this example is not to identify the optimal choice of containment measures, but to demonstrate the applicability of the PDMM to stochastic spreading processes and its efficiency on large real-world systems. One possible application is risk assessment for critical transitions of new diseases or virus variants between different countries or cities.

As a first approach to formulate the piecewise-deterministic spreading dynamics, we will consider the rate constants  $\gamma$ , which define the frequency of status changes for the population members, to be independent of the evolution of the process. Later, we will generalize the dynamics by letting these rates depend on time and on the history of the process.

### The System State

At the local scale of each subpopulation, we use a compartmental Susceptible-Exposed-Infected-Recovered-Deceased (SEIRD) model [39, 170] to describe the internal transmission dynamics of COVID-19. In this model, we consider five compartments for the infection status of individuals: susceptible (S), exposed (E), infected (I), recovered (R) and deceased (D). Susceptible individuals are the ones who have not yet been in contact with the virus and have no immunity against it. After being exposed to the virus, a susceptible individual is first in an asymptomatic status E that changes to a symptomatic status I after the end of an incubation period. Recovered and deceased individuals are removed from the transmission dynamics as we consider recovered individuals to be immune for the time scale of our model. Given these statuses as well as a set of  $m$  subpopulations, we can write the state of the system as

$$N = \left( N_S^{(k)}, N_E^{(k)}, N_I^{(k)}, N_R^{(k)}, N_D^{(k)} \right)_{k=1, \dots, m}$$

where  $N_i^{(k)}$  denotes the number of individuals in status  $i \in \{S, E, I, R, D\}$  within subpopulation  $k$ .

### Deterministic Local Interaction Dynamics

We assume that exposed individuals are already capable of transmitting the virus [171], so the status change from  $S$  to  $E$  can be caused by the second-order interactions of type either  $S + E \rightarrow 2E$  or  $S + I \rightarrow E + I$ , with corresponding rate constants  $\gamma_{SE}^{(k)} > 0$  and  $\gamma_{SI}^{(k)} > 0$ , see Figure 3.24 for an illustration. In general, it is possible to distinguish between infection rates from a contact with an exposed individual and from a contact with an infected individual. However, for simplicity, we assume here that infectivity is constant from the moment of exposure until recovery or death, i.e.,  $\gamma_{SE}^{(k)} = \gamma_{SI}^{(k)}$ . Given the system state  $N$ , the rate functions for the second-order status changes are defined by

$$f_{SE}^{(k)}(N) = \gamma_{SE}^{(k)} N_E^{(k)} N_S^{(k)}$$

and

$$f_{SI}^{(k)}(N) = \gamma_{SI}^{(k)} N_I^{(k)} N_S^{(k)},$$

respectively, compare (2.20). The remaining status changes that we consider are given by first-order events of the form  $E \rightarrow I$ ,  $I \rightarrow R$  and  $I \rightarrow D$  with respective rate constants  $\gamma_{EI}^{(k)}, \gamma_{IR}^{(k)}, \gamma_{ID}^{(k)} > 0$ . For the status change  $E \rightarrow I$  we accordingly obtain

$$f_{EI}^{(k)}(N) = \gamma_{EI}^{(k)} \cdot N_E^{(k)},$$

and equivalently for  $I \rightarrow R$  and  $I \rightarrow D$

$$f_{IR}^{(k)}(N) = \gamma_{IR}^{(k)} \cdot N_I^{(k)}, \quad f_{ID}^{(k)}(N) = \gamma_{ID}^{(k)} \cdot N_I^{(k)}.$$

The resulting ODE-system describing the local interaction dynamics within a subpopulation  $k$  is then given by

$$\begin{aligned} \frac{d}{dt} N_S^{(k)} &= - \left( \gamma_{SE}^{(k)} N_E^{(k)} + \gamma_{SI}^{(k)} N_I^{(k)} \right) N_S^{(k)} \\ \frac{d}{dt} N_E^{(k)} &= \left( \gamma_{SE}^{(k)} N_E^{(k)} + \gamma_{SI}^{(k)} N_I^{(k)} \right) N_S^{(k)} - \gamma_{EI}^{(k)} N_E^{(k)} \\ \frac{d}{dt} N_I^{(k)} &= \gamma_{EI}^{(k)} N_E^{(k)} - \left( \gamma_{IR}^{(k)} + \gamma_{ID}^{(k)} \right) N_I^{(k)} \\ \frac{d}{dt} N_R^{(k)} &= \gamma_{IR}^{(k)} N_I^{(k)} \\ \frac{d}{dt} N_D^{(k)} &= \gamma_{ID}^{(k)} N_I^{(k)}. \end{aligned} \tag{3.12}$$

### Stochastic Dynamics for Spatial Exchange

While the local interaction dynamics are given by deterministic evolution Equations (3.12), the spatial transitions between the subpopulations are described by stochastic

jump events. At any point in time  $t \geq 0$ , a member of subpopulation  $k$  can move to another subpopulation  $l \neq k$ , which induces a discrete change in the system state of the form

$$\mathbf{N}(t) \mapsto \mathbf{N}(t) - E_i^{(k)} + E_i^{(l)}$$

depending on the individual's status  $i \in \{S, E, I, R\}$ . (For  $i = D$  we naturally assume that jumps cannot take place.) The PDMM process combines these discrete, stochastic jump events between the subpopulations with the ODE dynamics (3.12) for local status-changes, which results in a stochastic process  $\mathbf{N}(t)_{t \geq 0}$ ,

$$\mathbf{N}(t) = \left( \mathbf{N}_S^{(k)}(t), \mathbf{N}_E^{(k)}(t), \mathbf{N}_I^{(k)}(t), \mathbf{N}_R^{(k)}(t), \mathbf{N}_D^{(k)}(t) \right)_{k=1, \dots, m},$$

described by an equation of the form (2.26). The terms in the first line of (2.26) thereby correspond to an integrated version of the ODE (3.12) for each  $k$ , and the second line describes the spatial jumps for given rate constants  $\lambda_i^{(kl)}$  between subpopulations  $k$  and  $l$ . In the following subsection, both the rate constants  $\gamma$  for the local interactions and the rate constants  $\lambda$  for the spatial transitions will depend on the evolution of the overall stochastic process  $\mathbf{N}(t)_{t \geq 0}$ .

**Remark 4.** Instead of physical transitions between subpopulations one can alternatively define adoption rate functions for rare in-between interactions. This is more appropriate if the subpopulations are defined in a social rather than a spatial sense, e.g., different age groups or different activities. The system of equations could then be modified accordingly, e.g., the process for transitions from S to E could be written as

$$\begin{aligned} d\mathbf{N}_S^{(k)} = & - \left( \gamma_{SE}^{(k)} \mathbf{N}_E^{(k)} + \gamma_{SI}^{(k)} \mathbf{N}_I^{(k)} \right) \mathbf{N}_S^{(k)} dt \\ & - \sum_{\substack{l=1 \\ l \neq k}} (E_E^k - E_S^{(k)}) (\gamma_{SE}^{(lk)} \mathbf{N}_E^{(l)} + \gamma_{SI}^{(lk)} \mathbf{N}_I^{(l)}) \mathbf{N}_S^{(k)} d\mathcal{P}_E^{(kl)}(t) \end{aligned}$$

with  $\gamma_{SE}^{(lk)} := \lambda_E^{(lk)} \gamma_{SE}^{(k)}$  being the rate for rare interactions between susceptible individuals of subpopulation  $l$  and exposed individuals of subpopulation  $k$  and  $\gamma_{SI}^{(lk)} := \lambda_E^{(lk)} \gamma_{SI}^{(k)}$  being the rate for rare interactions between susceptible individuals of subpopulation  $l$  and infected individuals of subpopulation  $k$  and  $\mathcal{P}_E^{(kl)}$  independent unit rate Poisson processes.

**Remark 5.** The modeling choices presented above are made for simplicity and to demonstrate how the PDMM can be used to conceptually analyze COVID-19 spreading. With rich data sets and extensive literature on COVID-19 pandemics, our model can be easily extended to include more realistic scenarios. For example, considering additional compartments such as symptomatic, asymptomatic, quarantined individuals [172]; including more general infection rates with possible time dependence [173], adding demographic information [174, 175], introducing vaccination effects [174] are just some of the extensions that would make this model more realistic.

## Adaptive regulation of rate constants

At the beginning of the COVID-19 pandemic, much effort was made to slow the spreading of the virus through non-pharmaceutical interventions, such as the introduction of measures to reduce social contact. This is achieved by targeting individual interactions (e.g. social distancing and wearing masks), reducing the number of interactions (e.g. closing schools, offices), but also by introducing global measures such as travel bans between countries and continents. In this sense, the choice of transition rate constants  $\gamma$  that are independent of time and the evolution of the process seems unrealistic, since containment measures are taken depending on the dynamics in order to influence the future evolution of the process. Therefore, in the following we will consider rate constants that are adapted over time according to given rules.

Transmission rates depend on the local contacts within a population, which change over time as interventions are implemented. In addition, the dependence may also depend on the type of interaction, e.g., exposed individuals may be less infectious than infected ones, or symptomatic cases may cause fewer infections than expected because they have already reduced their number of contacts. To accommodate many types of possible dependencies, we define transmission rates in a very general way as functions of the process history  $N_{\leq t} := (N(s))_{s \leq t}$  and time  $t$

$$\gamma_{SE}^{(k)} = \gamma_{SE}^{(k)}(N_{\leq t}, t), \quad \gamma_{SI}^{(k)} = \gamma_{SI}^{(k)}(N_{\leq t}, t).$$

This means that these rates depend not only on the current state, but also on the history of the process. This allows us to define rules, such as implementing a strict lockdown when the number of cases increases for the first time. In our model, the rate for developing symptoms  $\gamma_{EI} \geq 0$  is a constant, while the recovery and case fatality rates depend on the capacity of a population's health care system, so they are defined as state-dependent rates  $\gamma_{IR}^{(k)} = \gamma_{IR}^{(k)}(N)$  and  $\gamma_{ID}^{(k)} = \gamma_{ID}^{(k)}(N)$ . One of the goals of the measures introduced is to control the number of infections so that the limits of the health care system are not reached. As part of the global measures, the transitions between the subpopulations are reduced. Thus, the spatial transition rates  $\lambda_i^{(kl)}$  between the subpopulations are defined as functions of the history and time of the entire process for  $i \in \{S, E, I, R\}$ :

$$\lambda_i^{(kl)} = \lambda_i^{(kl)}(N_{\leq t}, t).$$

## Concrete Choice of Rate Constants for Status Changes

When modeling the implementation of virus containment measures, we assume that each measure is followed by a phase in which the infection rate remains constant. The transition between phases can be triggered by either deterministic or stochastic events, such as the process crossing a threshold number of infections for the first time. In total, we consider three different phases:

1. **Initial phase:** At the beginning of the pandemic, the infection rates  $\gamma_{SE}^{(k)}$  and  $\gamma_{SI}^{(k)}$  have values  $\delta_{SE} > 0$  and  $\delta_{SI} > 0$ , respectively, and the interaction dynamics start with unrestricted spreading.
2. **Strict measures phase:** The first measures to reduce the infection rates are taken in subpopulation  $k$  as soon as the number of infected individuals exceeds a critical value  $h_I^{(k)} > 0$  for the first time. That is, the strict measures start at the random first hitting time

$$t_1^{(k)}(N_{\leq t}) := \min \left\{ 0 \leq s \leq t \mid \mathbf{N}_I^{(k)}(s) \geq h_I^{(k)} \right\} \in [0, \infty],$$

with value  $t_1^{(k)}(N_{\leq t}) = \infty$  in case of  $N_I^{(k)}(s) < h_I^{(k)}$  for all  $s \leq t$ . In this phase, the infection rates  $\gamma_{SE}^{(k)}$  and  $\gamma_{SI}^{(k)}$  are reduced by a factor  $\kappa_1^{(k)} \in (0, 1)$ . These strict measures are maintained until the number of infected individuals falls below the critical value  $\frac{h_I^{(k)}}{2}$ , i.e., until the random time point

$$t_2^{(k)}(N_{\leq t}) := \min \left\{ t_1^{(k)} < s \leq t \mid \mathbf{N}_I^{(k)}(s) < \frac{h_I^{(k)}}{2} \right\} \in [0, \infty].$$

3. **Moderate measures phase:** After the strict measures are lifted, the interactions within the population do not return to normal, i.e., to the values from the initial phase. Instead, we introduce moderate measures where the infection rates are scaled with a factor  $\kappa_2^{(k)}$ , s.t.  $\kappa_1^{(k)} < \kappa_2^{(k)} < 1$ , allowing for more contacts than in the previous phase. These measures are maintained for the remaining time of the model even if the number of infected individuals again crosses the value  $h_I^{(k)}$ .

Taken all together, this means that the infection rate function is defined by

$$\gamma_{SE}^{(k)}(N_{\leq t}, t) := \begin{cases} \delta_{SE} & \text{for } t \leq t_1^{(k)}(N_{\leq t}) \\ \kappa_1^{(k)} \delta_{SE} & \text{for } t_1^{(k)}(N_{\leq t}) < t \leq t_2^{(k)}(N_{\leq t}) \\ \kappa_2^{(k)} \delta_{SE} & \text{for } t_2^{(k)}(N_{\leq t}) < t \end{cases} \quad (3.13)$$

and equivalently for  $\gamma_{SI}^{(k)}$ , possibly with different reduction factors  $\kappa_i^{(k)}$   $i = 1, 2$ . More generally, also the rate constants  $\delta_{SE}$  can depend on the subpopulation  $k$ , but here we omit the corresponding indices for the purpose of simplicity.

In order to make our model more realistic, we include in each subpopulation  $k$  a limited health care capacity given by a threshold  $h_R^{(k)}$ . We assume that the case fatality rate  $\gamma_{ID}^{(k)} \geq 0$  increases from a given value  $\delta_{ID}$  to another value  $\tilde{\delta}_{ID} > \delta_{ID}$  if the number of infected individuals exceeds this threshold  $h_R^{(k)}$ , giving

$$\gamma_{ID}^{(k)}(N) := \begin{cases} \delta_{ID} & \text{for } N_I^{(k)} \leq h_R^{(k)} \\ \tilde{\delta}_{ID} & \text{for } N_I^{(k)} > h_R^{(k)}. \end{cases}$$

Vice versa, the recovery rate  $\gamma_{IR}$  is reduced in case of an exhausted health care capacity, such that

$$\gamma_{IR}^{(k)}(N) := \begin{cases} \delta_{IR} & \text{for } N_I^{(k)} \leq h_R^{(k)} \\ \tilde{\delta}_{IR} & \text{for } N_I^{(k)} > h_R^{(k)} \end{cases}$$

for constants  $\delta_{IR} > \tilde{\delta}_{IR} \geq 0$ . Additionally, within each subpopulation  $k$  we consider  $\gamma_{ID}^{(k)} + \gamma_{IR}^{(k)}$  to be constant, i.e.,  $\delta_{IR} + \delta_{ID} = \tilde{\delta}_{IR} + \tilde{\delta}_{ID}$ .

Finally, we assume that exposed individuals develop symptoms after an incubation period of average length  $\tau_{EI} > 0$  and set

$$\gamma_{EI}^{(k)} = \frac{1}{\tau_{EI}}$$

for all  $k$ .

### Concrete Choice of Rate Constants for Spatial Transitions

The global spatial transition rate functions between the subpopulations  $\lambda_i^{(kl)}$  will depend on the local phases within each of the subpopulations. More precisely, we define

$$\tau_1(N_{\leq t}) := \min \left\{ t_1^{(1)}(N_{\leq t}), t_1^{(2)}(N_{\leq t}) \right\}$$

to be the first time that one of the subpopulations initiates the lock-down phase and

$$\tau_2(N_{\leq t}) := \max \left\{ t_2^{(1)}(N_{\leq t}), t_2^{(2)}(N_{\leq t}) \right\}$$

to be the first time that both subpopulations have ended the lock-down phase. Whenever in one of the subpopulations the strict measures are applied, the spatial transition rates are reduced by a factor  $\kappa_1^{(kl)} \in (0, 1)$ . After the strict measures have ended in both populations, the spatial transition rates are scaled by a factor  $\kappa_2^{(kl)}$ , where  $\kappa_1^{(kl)} < \kappa_2^{(kl)} < 1$ . Thus, the spatial transition rate functions are defined as

$$\lambda_i^{(kl)}(N_{\leq t}, t) := \begin{cases} \delta^{(kl)}, & \text{for } t \leq \tau_1(N_{\leq t}) \\ \kappa_1^{(kl)} \delta^{(kl)} & \text{for } \tau_1(N_{\leq t}) < t \leq \tau_2(N_{\leq t}) \\ \kappa_2^{(kl)} \delta^{(kl)} & \text{for } \tau_2(N_{\leq t}) < t \end{cases} \quad (3.14)$$

for  $i \in \{S, E, R\}$ . We assume that people with symptoms do not travel, i.e., we set  $\lambda_I^{(kl)} = \lambda_D^{(kl)} = 0$  for all  $k, l$  independently of time.

### PDMM-based simulations of COVID-19 spreading

We simulate the dynamics for model scenarios with different infection and spatial transition dynamics. In particular, we compare the following three scenarios:

**Scenario 1:** Choose constant infection and spatial transition rates that can be interpreted as no action being taken.

**Scenario 2:** Let the infection rate depend on the process history as defined in Equation (3.13), but assume constant spatial transition rates between subpopulations. This corresponds to introducing local measures to control the infection dynamics within subpopulations, but no additional travel restrictions between them.

**Scenario 3:** Combine the measures, i.e., let both infection and spatial transition rates change according to the epidemic dynamics following the rules defined in (3.13) and (3.14).

### Parameter Choices

Recently, much research has been done to infer the parameters of the COVID-19 dynamics from available data. However, for most parameters there is a wide range of estimates, and thus the choice for a conceptual model can seem arbitrary. Since parameter estimation is not the focus of this work, but the modeling approach is, we will choose the parameters based on several recent publications [176–180].

**Parameters for status change  $E \rightarrow I$ :** The average incubation period was estimated to be 5 – 6 days [178, 179], so we choose  $\tau_{EI} = 5.5$ .

**Parameters for status change  $I \rightarrow R$  and  $I \rightarrow D$ :** The average time for transition from I to either R or D will be set to 14 days [178, 179]. For the infection fatality rate of our model, we will use the estimate from [180], which leads to the choices of  $\delta_{IR} = \frac{1-0.014}{14}$  and  $\delta_{ID} = \frac{0.014}{14}$  for the recovery and case fatality rates of the populations. We choose  $\tilde{\delta}_{ID} = 3\delta_{ID}$  and  $\tilde{\delta}_{IR}$  accordingly such that  $\delta_{IR} + \delta_{ID} = \tilde{\delta}_{IR} + \tilde{\delta}_{ID}$  is fulfilled.

**Parameters for status change  $S \rightarrow E$ :** Here we consider two reactions that can lead to the status change  $S \rightarrow E$ , namely  $S + E \rightarrow 2E$  and  $S + I \rightarrow E + I$ . As discussed before, for simplicity, we assume that  $\gamma_{SE}^{(k)} = \gamma_{SI}^{(k)}$  and thus  $\delta_{SI} = \delta_{SE}$ . Estimates for the initial reproduction number  $R_0 := \frac{\delta_{SE}}{\delta_{IR} + \delta_{ID}}$  vary depending on the region of choice [181] as well as on the estimation method [182]. This leads to a wide range of possible parameter choices that have the highest impact on the model outcome. For our model, we use  $R_0 = 4.1$  which corresponds to the estimate for the New York City in [181], assuming our subpopulations to be well-mixed and an urban area like NYC to meet this assumption. This choice of  $R_0$  leads to the infection rate  $\delta_{SE} := \frac{4.1}{14}$ . The remaining parameters will be subject to changes due to different containment measures. During a period of strict measures phase we will reduce the infection rate to 10% of the original value by setting a scaling factor  $\kappa_1^{(k)} = 0.1$  for  $k = 1, 2$ . In the phase of moderate measures, we assume more interactions which lead to an increase of the infection rate. For illustration purposes, we consider different choices of moderate measures in each subpopulation, such that we set values of the infection rates to be 30% and 40% of

the original  $\delta_{SE}$ , i.e.,  $\kappa_2^{(1)} = 0.3$  and  $\kappa_2^{(2)} = 0.4$ . For both subpopulations, the infection threshold for the first hitting time event is chosen to be 2% of the initial total population number  $n_a^{(k)}$  in subpopulation  $k$ , i.e.,  $h_I^{(k)} = 0.02 \cdot n_a$  and the threshold for the capacity of the health care system is reached when 10% of  $n_a$  are infected, i.e.,  $h_R^{(k)} = 0.1 \cdot n_a^{(k)}$ .

**Parameters for spatial transitions:** The spatial transition rates are chosen to be  $\delta^{(kl)} = \delta^{(lk)} = 0.0003$ , which corresponds in our example to 3 out of 10000 agents transitioning per time unit and fits to the assumption of a metastable setting with slow transitions between subpopulations compared to the infection dynamics within subpopulations. When at least one subpopulation is in the strict measures phase, we introduce travel restrictions by reducing the spatial transition rate to 5% of the original value, i.e., we set  $\kappa_1^{kl} = \kappa_1^{lk} = 0.05$ . When both subpopulations are in the moderate measures phase, we moderately relax the travel restrictions by increasing the spatial transition rate to 50% of the original value, i.e., we set  $\kappa_2^{kl} = \kappa_2^{lk} = 0.5$ .

### Simulation Results

For the initial population sizes we choose the values  $n_a^{(1)} = n_a^{(2)} = 10\,000$ . We start with one member of the subpopulation 1 (SP1) being in status E and all other members of SP1 and subpopulation 2 (SP2) being in status S. The critical transition event is the first time when an individual with status E jumps from SP1 to SP2. In Figure 3.25, we

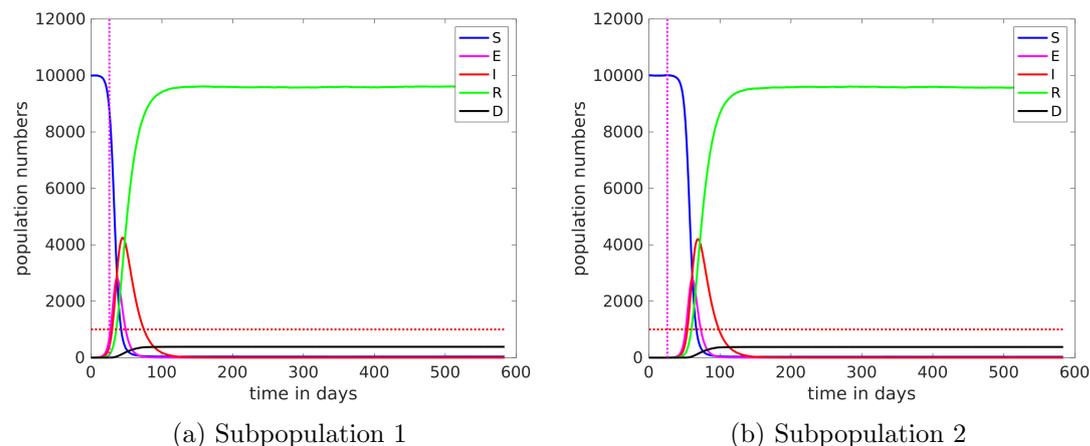


Figure 3.25: Trajectory of a PDMM simulation for Scenario 1 (no measures). The dotted magenta line marks the critical transition event, the horizontal red line marks the threshold  $h_R$  of health care capacity.

see one outcome of Scenario 1, where no containment measures are taken. We observe one wave of infections in each subpopulation with the number of I cases quickly rising above the threshold  $h_R$  and staying there for a considerable amount of time. Almost all members of the population get infected. Due to the increased case fatality rate  $\gamma_{ID}$ , at

the end 3.8% of the total population is in status D. This is a scenario that should be avoided in reality, e.g., by flattening the curve by implementing containment measures.

As a result of the local measures that are present in Scenario 2, the infection curve shows two smaller waves instead of one large wave, see Figure 3.26. In SP1, the number of infections remains below the threshold  $h_R$  throughout the simulation period, while in SP2 the number of cases exceeds  $h_R$  during the peak of the second wave. This is due to a higher infection rate within SP2 in the phase of moderate measures, which leads to a higher total number of infections and more fatal cases in SP2 at the end of the simulation. Nevertheless, the outcome in both subpopulations is a much smaller number of I and D individuals than in Scenario 1. The same is true for Scenario 3 which has the same local interventions. In addition, the shape of the infection curves determined

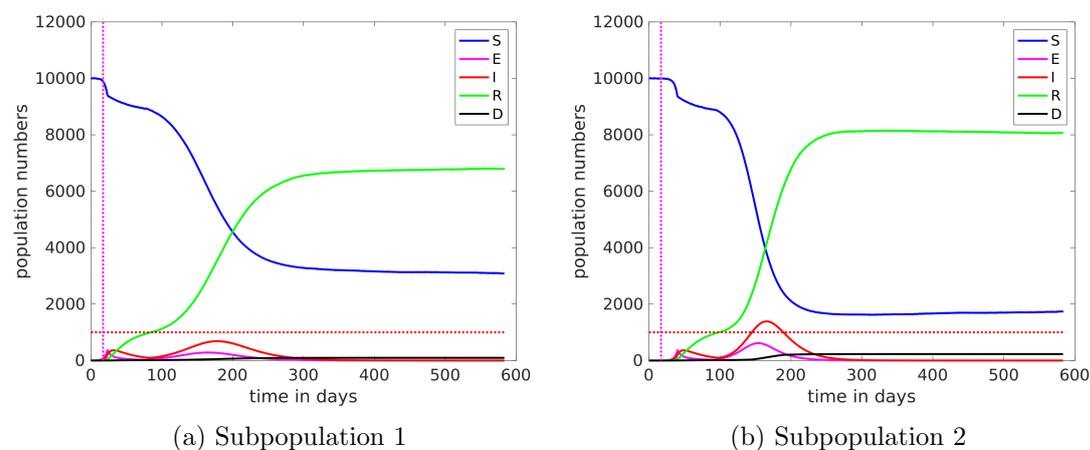


Figure 3.26: Trajectory of a PDMM simulation for Scenario 2. The dotted magenta line marks the critical transition event, the horizontal red line marks the threshold  $h_R$ .

by the internal population dynamics is the same in Scenarios 2 and 3 (see Figure 3.27). However, the distribution of the critical transition time that starts the epidemic in SP2 is significantly different. Due to the introduction of travel restrictions in Scenario 3, we observe a later first infection in SP2 compared to the one from Scenario 2, see Figure 3.27b.

In order to compare the critical transition time distributions for different containment measures, we run 10000 MC-simulations for each of the three scenarios, see Figure 3.28. For about a third of the simulations (regardless of the scenario), the critical transition occurs before time  $t_1^{(1)}$  when the measures are introduced in scenarios 2 and 3. After this time, we can observe the differences in the shapes of the critical transition time distributions. Namely, compared to Scenario 1, we observe for Scenario 2 a larger number of critical transitions occurring later in time. This is due to the influence of the number

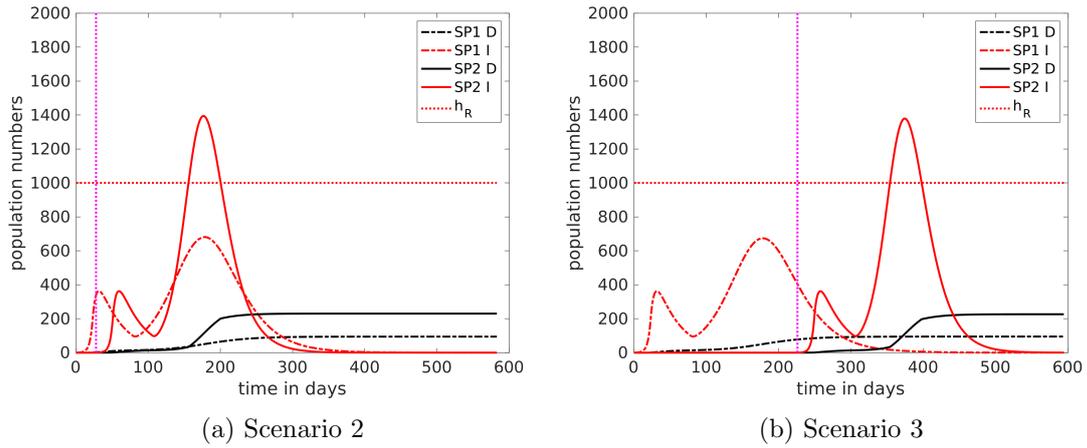


Figure 3.27: Comparison between infection curves for single trajectories obtained in Scenarios 2 and 3. The vertical dotted magenta line marks the critical transition event, the horizontal red line marks the threshold  $h_R$ . The dashed lines refer to the development in the subpopulation 1 and solid lines to subpopulation 2.

of active cases in SP1, which decreases much faster in Scenario 2 than in Scenario 1 due to the local interventions. The mean time for the first infection in SP2 is 24.5 days for Scenario 1 and 43.9 days for Scenario 2. Due to the reduced number of spatial transitions in Scenario 3, the probability of a critical transition during the measures is much lower. As a result, in 3989 (out of 10000) MC-simulations the critical transition event did not happen at all, and the virus was successfully contained in SP1. Conditional on the transition occurring before the end of the simulation period, the mean critical transition time was 78.6 days, which demonstrates the benefit of the travel restrictions on the spreading dynamics.

In our conceptual model the outcome of the epidemic was improved by the implementation of measures. The spatial separation into multiple subpopulations has a strong impact on the dynamics of the spreading process, especially when considering the scenario of combined measures where the transmission between populations could be delayed for a long time and sometimes even be prevented. The model results highlight the importance of a rapid response to a new disease or virus variant, as the probability of spreading between separated populations is much higher before any measures are taken.

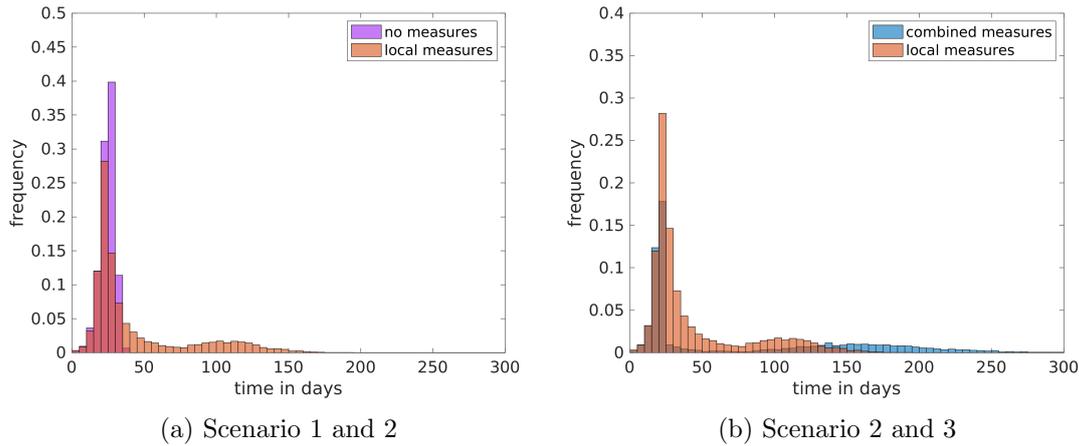


Figure 3.28: Critical transition time distribution for different scenarios of the model. The number of MC-simulations for each scenario is 10000.

### 3.3.3 Discussion of ABM Approaches and Model Reduction

While we are able to incorporate preventive measures into the macro- and mesoscale models of epidemic spreading, we can only do so if we already have an assessment of how the measures introduced affect the rate constants of the model. We are not able to gain insight into the effectiveness of specific interventions, such as school closures or mask use, beyond the information already built into the model assumptions. Questions like these can only be addressed by a microscale model with more realistic dynamics for the interactions of individuals within the population (or subpopulations).

An agent-based model for epidemic spreading on the microscale faces two major challenges compared to the ODE and metapopulation model approaches. The first challenge is to build a realistic model for the mobility of the agents that replaces the well-mixed assumption. The second challenge is the model parametrization as the microscopic rate constants, e.g., for transmission of the virus, are usually unknown and have to be estimated by fitting the model outcome to available data. Since ABM simulations are typically computationally expensive, this may not be feasible if the parameter space is too high dimensional.

#### Modeling of contemporary mobility

While we assumed in the models for prehistoric social spreading processes that the mobility of the agents was mainly characterized by foraging patterns, the mobility of the present day society can be vastly different. Though there still exist hunter-gatherers today, the majority of people are living in urban areas [183]. Within cities, individuals choose their location in general based on the activity that they pursue at the moment or in the near future. When a target destination is known, then in general also a shortest

or fastest path to it is known, as well as a preferred mode of transportation to get there. Exploratory movements are rather the exception and usually limited to the choice of travel destinations, e.g., going to a new place for a leisure activity. Once the activities are scheduled, the mobility decisions are mostly determined besides reactions to possible unknown factors such as traffic jams or delayed departure times due to unforeseen events. Therefore, to model these mobility dynamics we need a completely different approach than in the previous agent-based models.

State-of-the-art ABMs for mobility typically feature agents with a state variable consisting of a schedule of activities and their associated locations (home, office, school, etc.) [184, 185]. Each location has a geographical position and depending on the scope of the model the traffic resulting from agents commuting between the different locations is also explicitly modeled [184]. Especially in the context of epidemic modeling, it is desirable to include the possibility for agents to meet while using public transportation.

In the context of epidemic spreading, the agents also possess the status variables that are relevant for the research question, such as the epidemiological status (e.g., SEIR) or the age group. The transmission dynamics are usually realized by pairwise interactions, with agents being able to interact if they are at the same location (building or vehicle) [158, 185]. As the locations for activities are discrete, though they are linked to a location in continuous space, this type of ABM could also be interpreted as a large scale metapopulation model [186]. Many models consider the mobility and the epidemic dynamics to influence each other, the transmission dynamics clearly depends on the changes of the position and the schedule of the agents is adapted depending on the status, e.g., after infection or recovery events.

At this microscopic scale, different specific measures can be emulated and their impact on infection dynamics can be evaluated. Some measures can be implemented by adjusting the schedules of the agents. Typical examples are school closings, home offices, or isolation of infected agents [158, 185]. Wearing masks in certain locations, such as public transportation, can be implemented by adjusting the infection rate associated with the location. Many models also distinguish between indoor and outdoor activities and seasonal influences, and define different infection rates for each scenario. However, the more realistic scenarios are more difficult to parameterize.

### **Model reduction**

The spatial scale of ABMs or household models for epidemic spreading is usually involving at most one city. While the dynamics of mobility models based on real world data can in general not be considered well-mixed as spatial effects can be observed [187, 188], there often is also no clear timescale separation possible to identify metastable areas. Also, the interactions between different age groups, while less frequent than those within groups, do not occur on a significantly slower time scale. Thus, the model reduction approach from Chapter 2 is in general not applicable to these models as the existence of

metastable dynamics is crucial for a good approximation quality. Without metastability the model does not profit from the assumption of a heterogeneous instead of a well-mixed population and we can just apply a mean field approach for model reduction [189]. The resulting model is then a compartmental ODE model where the interactions of the agent population are approximated by well-mixed dynamics.

Since the model reduction process may aggregate multiple microscale parameters into a single macroscopic parameter (e.g., multiple microscopic infection rates), it may not be feasible to explicitly compute the macroscopic parameters from the microscopic parameters. Thus, for the construction of compartmental ODE models based on ABMs on the microscale, it is common to utilize data driven approaches such as SINDy [190] or PREDICI [158] for the derivation of the macroscopic parameters. For the reduced models it is then also possible to perform parameter optimization, e.g. by calculating the Pareto front [158].

# Conclusion

Agent-based models play a central role in modeling social spreading processes, in part because they allow detailed representation of interactions between individuals while integrating data on real-world processes. However, the resulting models are often too complex for a formal analysis and usually require high simulation effort. In this thesis, based on general remarks on theoretical concepts such as stochastic dynamics and Markov processes, we have first presented some new theoretical results on the efficient simulation and model reduction of agent-based models. Among these results are an event-based simulation algorithm for ABMs and a model reduction approach based on a projection of the state space and the utilization of convergence results to approximate agent-based models by less complex metapopulation models that can be simulated with much less effort. Assuming metastability of the agent system, this approach preserves important model characteristics with a low approximation error.

In relation to this background a number of applications of agent-based models have been discussed. Of these, some are of fundamental structure, including a model to achieve global goals with local information, and others concern concrete spreading processes in prehistoric and contemporary societies. A focus among the applications is the spreading and development of culture and innovations in ancient times, both on a conceptual level and with reference to a concrete application case, the spread of the woolly sheep to Europe. In this context, the presented models have been developed through interdisciplinary cooperation and by taking into account archaeological, anthropological as well as geographical data in order to be able to depict the mobility and interactions of nomads, such as hunter-gatherers or shepherds, as realistically as possible. An important aspect that was discussed is the challenges posed by the prehistoric context, both in model parameterization and in validation and interpretation of the results. The comparison with current modeling scenarios is discussed with reference to the application area of epidemic spreading. Specifically, the differences in the assumptions about agent mobility and in the availability and reproducibility of data relevant to the model construction and analysis are highlighted.

In the analysis of the models, we focused in particular on the identification of metastable processes through the application of clustering methods, including a novel approach that exploits the specific structure of the agent-based models we have presented. Based on this analysis, possibilities for model reduction were discussed, which allow to generate additional data on macroscopic properties and mesoscopic structures of the models with low effort. Especially, the generation of relevant statistics about critical transitions and other rare events is enabled by the reduced model complexity.

## Zusammenfassung

Agentenbasierte Modelle spielen bei der Modellierung sozialer Ausbreitungsprozesse eine zentrale Rolle, da sie unter anderem die Interaktionen zwischen Individuen detailliert abbilden und Daten über reale Prozesse integrieren können. Die resultierenden Modelle sind jedoch häufig zu komplex für eine formale Analyse und in der Regel mit einem hohen Simulationsaufwand verbunden. In dieser Arbeit werden zunächst, aufbauend auf allgemeinen Ausführungen zu theoretischen Konzepten wie stochastischer Dynamik und Markov-Prozessen, einige neue theoretische Ergebnisse zur Simulation und Modellreduktion von agentenbasierten Modellen vorgestellt. Hervorzuheben ist dabei ein auf einer Zustandsraumprojektion basierender Ansatz zur Approximation agentenbasierter Modelle durch weniger komplexe Metapopulationsmodelle, die mit deutlich geringerem Aufwand simuliert werden können. Unter der Voraussetzung der Metastabilität bleiben dabei wichtige Modelleigenschaften bei geringem Approximationsfehler erhalten.

In diesem Zusammenhang und im Anschluss daran werden eine Reihe von Anwendungen agentenbasierter Modelle diskutiert. Einige davon sind grundlegender Natur, darunter ein Modell zur Erreichung globaler Ziele mit lokalen Informationen, andere betreffen konkrete Ausbreitungsprozesse in prähistorischen und zeitgenössischen Gesellschaften. Ein Schwerpunkt unter den Anwendungsbereichen ist die Ausbreitung und Entwicklung von Kultur und Innovationen in der Antike, sowohl auf konzeptioneller Ebene als auch in Bezug auf einen konkreten Anwendungsfall, die Ausbreitung des Wollschafs nach Europa. Dabei wurden die vorgestellten Modelle in interdisziplinärer Kooperation und unter Berücksichtigung archäologischer, anthropologischer und geographischer Daten entwickelt, um die Mobilität und Interaktionen von Nomaden wie Jägern und Sammlern oder Hirten möglichst realitätsnah abbilden zu können. Ein wichtiger Aspekt, der diskutiert wird, sind die Herausforderungen, die sich aus dem prähistorischen Kontext sowohl für die Modellparametrisierung als auch für die Validierung und Interpretation der Ergebnisse ergeben. Der Vergleich mit aktuellen Modellierungsszenarien wird in Bezug auf das Anwendungsgebiet der Infektionsausbreitung diskutiert. Dabei werden insbesondere die Unterschiede in den Annahmen zur Mobilität der Agenten und in der Verfügbarkeit und Reproduzierbarkeit der für die Modellkonstruktion und -analyse relevanten Daten hervorgehoben.

Bei der Analyse der Modelle liegt ein besonderer Schwerpunkt auf der Identifikation metastabiler Prozesse durch die Anwendung von Clusterverfahren, einschließlich eines neuartigen Ansatzes, der die besondere Struktur agentenbasierter Modelle ausnutzt. Darauf aufbauend werden Möglichkeiten der Modellreduktion diskutiert, die es erlauben, mit geringem Aufwand zusätzliche Daten über makroskopische Eigenschaften und mesoskopische Strukturen der Modelle zu erzeugen. Insbesondere die Generierung relevanter Statistiken über kritische Übergänge und andere seltene Ereignisse wird durch die geringere Modellkomplexität erst ermöglicht.

# Bibliography

- [1] Elia Gabarron, Sunday Oluwafemi Oyeyemi, and Rolf Wynn. Covid-19-related misinformation on social media: a systematic review. *Bulletin of the World Health Organization*, 99(6):455, 2021.
- [2] Sebastian Alexander Müller, William Charlton, Natasa Djurdjevac Conrad, Ricardo Ewert, Sydney Paltra, Christian Rakow, Tim Conrad, Christof Schütte, and Kai Nagel. Modus-covid bericht vom 23.02. 2022. 2022.
- [3] Adam J McLane, Christina Semeniuk, Gregory J McDermid, and Danielle J Marceau. The role of agent-based models in wildlife ecology and management. *Ecological modelling*, 222(8):1544–1556, 2011.
- [4] Nigel Gilbert and Pietro Terna. How to build and use agent-based models in social science. *Mind & Society*, 1(1):57–72, 2000.
- [5] Leigh Tesfatsion. Agent-based computational economics: Growing economies from the bottom up. *Artificial life*, 8(1):55–82, 2002.
- [6] Elizabeth Hunter, Brian Mac Namee, and John D Kelleher. A taxonomy for agent-based models in human infectious disease epidemiology. *Journal of Artificial Societies and Social Simulation*, 20(3), 2017.
- [7] Iza Romanowska, Colin D Wren, and Stefani A Crabtree. *Agent-based modeling for archaeology: Simulating the complexity of societies*. SFI Press, 2021.
- [8] Fabian Becker, Nataša Djurdjevac Conrad, Raphael A Eser, Luzie Helfmann, Brigitta Schütt, Christof Schütte, and Johannes Zonker. The furnace and the goat—a spatio-temporal model of the fuelwood requirement for iron metallurgy on elba island, 4th century bce to 2nd century ce. *PloS one*, 15(11):e0241133, 2020.
- [9] Frank Schweitzer and J Doyne Farmer. *Brownian agents and active particles: collective dynamics in the natural and social sciences*, volume 1. Springer, 2003.
- [10] Sven Banisch. *Markov chain aggregation for agent-based models*. Springer, 2015.
- [11] Akira Namatame and Shu-Heng Chen. *Agent-based modeling and network dynamics*. Oxford University Press, 2016.
- [12] Daniel T Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.

- [13] Christof Schütte and Marco Sarich. *Metastability and Markov State Models in Molecular Dynamics*, volume 24 of *Courant Lecture Notes*. American Mathematical Soc., 2013.
- [14] Natasa Djurdjevac Conrad, Daniel Fuerstenau, Ana Grabundzija, Luzie Helfmann, Martin Park, Wolfram Schier, Brigitta Schütt, Christof Schütte, Marcus Weber, Niklas Wulkow, and Johannes Zonker. Mathematical modeling of the spreading of innovations in the ancient world. *eTopoi. Journal for Ancient Studies*, 7, 2018.
- [15] Nataša Djurdjevac Conrad, Luzie Helfmann, Johannes Zonker, Stefanie Winkelmann, and Christof Schütte. Human mobility and innovation spreading in ancient times: a stochastic agent-based simulation approach. *EPJ Data Science*, 7(1):1–22, 2018.
- [16] 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.
- [17] 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.
- [18] Peter Van Inwagen. The incompatibility of free will and determinism. *Philosophical studies*, 27(3):185–199, 1975.
- [19] Joseph L Doob. *Stochastic processes*, volume 7. Wiley New York, 1953.
- [20] Philip E Protter. Stochastic differential equations. In *Stochastic integration and differential equations*, pages 249–361. Springer, 2005.
- [21] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.
- [22] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.
- [23] Achim Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.
- [24] Martin Hairer. Ergodic properties of markov processes. *Lecture notes*, 2006.
- [25] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.
- [26] Alexander Grigor’Yan. Analytic and geometric background of recurrence and non-explosion of the brownian motion on riemannian manifolds. *Bulletin of the American Mathematical Society*, 36(2):135–249, 1999.

- [27] Andreas E Kyprianou. *Introductory lectures on fluctuations of Lévy processes with applications*. Springer Science & Business Media, 2006.
- [28] David Applebaum. *Lévy processes and stochastic calculus*. Cambridge university press, 2009.
- [29] Mark H A Davis. Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 46(3):353–388, 1984.
- [30] Uwe Franz, Volkmar Liebscher, and Stefan Zeiser. Piecewise-deterministic Markov processes as limits of Markov jump processes. *Advances in Applied Probability*, 44(3):729–748, 2012.
- [31] Thomas G Kurtz. Limit theorems and diffusion approximations for density dependent markov chains. In *Stochastic Systems: Modeling, Identification and Optimization, I*, pages 67–78. Springer, 1976.
- [32] José E Figueroa-López. Jump-diffusion models driven by lévy processes. In *Handbook of Computational Finance*, pages 61–88. Springer, 2012.
- [33] George EP Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, 1976.
- [34] Craig W Reynolds. Flocks, herds and schools: A distributed behavioral model. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 25–34, 1987.
- [35] Frank Schweitzer. *Brownian agents and active particles: collective dynamics in the natural and social sciences*. Springer, 2007.
- [36] Robert Axelrod. Agent-based modeling as a bridge between disciplines. *Handbook of computational economics*, 2:1565–1584, 2006.
- [37] Wenwu Tang and David A Bennett. Agent-based modeling of animal movement: a review. *Geography Compass*, 4(7):682–700, 2010.
- [38] Zhihui Wang, Joseph D Butner, Romica Kerketta, Vittorio Cristini, and Thomas S Deisboeck. Simulating cancer growth with multiscale agent-based modeling. In *Seminars in cancer biology*, volume 30, pages 70–78. Elsevier, 2015.
- [39] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.
- [40] Flavio Iannelli, Andreas Koher, Dirk Brockmann, Philipp Hövel, and Igor M Sokolov. Effective distances for epidemics spreading on complex networks. *Physical Review E*, 95(1):012313, 2017.

- [41] Jan Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12):1819–1838, 2007.
- [42] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, Seong Joon Kim, and Song Chong. On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking*, 19(3):630–643, 2011.
- [43] Hugo Barbosa, Marc Barthélemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.
- [44] Arnaud Debussche, Michael Högele, and Peter Imkeller. *The dynamics of nonlinear reaction-diffusion equations with small Lévy noise*, volume 2085. Springer, 2013.
- [45] Masao Doi. Stochastic theory of diffusion-controlled reaction. *Journal of Physics A: Mathematical and General*, 9(9):1479, 1976.
- [46] Charles M Macal and Michael J North. Agent-based modeling and simulation. In *Proceedings of the 2009 winter simulation conference (WSC)*, pages 86–98. IEEE, 2009.
- [47] Yimin Chen, Xia Li, Xiaoping Liu, and Yilun Liu. An agent-based model for optimal land allocation (agentla) with a contiguity constraint. *International Journal of Geographical Information Science*, 24(8):1269–1288, 2010.
- [48] René Jordan, Mark Birkin, and Andrew Evans. An agent-based model of residential mobility: Assessing the impacts of urban regeneration policy in the easel district. *Computers, Environment and Urban Systems*, 48:49–63, 2014.
- [49] Nataša Djurdjevac Conrad, Jonas Köppl, and Ana Djurdjevac. Feedback loops in opinion dynamics of agent-based models with multiplicative noise. *Entropy*, 24(10):1352, 2022.
- [50] Mark Pogson, Mike Holcombe, Rod Smallwood, and Eva Qwarnstrom. Introducing spatial information into predictive nf- $\kappa$ b modelling—an agent-based approach. *PLoS One*, 3(6):e2367, 2008.
- [51] Chad M Glen, Melissa L Kemp, and Eberhard O Voit. Agent-based modeling of morphogenetic systems: Advantages and challenges. *PLoS computational biology*, 15(3):e1006577, 2019.
- [52] Jeremy BA Green and James Sharpe. Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development*, 142(7):1203–1211, 2015.
- [53] Peter E Kloeden and Eckhard Platen. Stochastic differential equations. In *Numerical solution of stochastic differential equations*, pages 103–160. Springer, 1992.

- [54] Benedict Leimkuhler and Charles Matthews. Rational construction of stochastic numerical methods for molecular sampling. *Applied Mathematics Research eXpress*, 2013(1):34–56, 2013.
- [55] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.
- [56] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of chemical physics*, 115(4):1716–1733, 2001.
- [57] Yang Cao, Daniel T Gillespie, and Linda R Petzold. Avoiding negative populations in explicit poisson tau-leaping. *The Journal of chemical physics*, 123(5):054104, 2005.
- [58] Christian L Vestergaard and Mathieu Géniois. Temporal Gillespie algorithm: Fast simulation of contagion processes on time-varying networks. *PLoS computational biology*, 11(10):e1004579, 2015.
- [59] Yucheng Hu, Tiejun Li, and Bin Min. A weak second order tau-leaping method for chemical kinetic systems. *The Journal of chemical physics*, 135(2):024113, 2011.
- [60] Graeme D Chalmers and Desmond J Higham. Asymptotic stability of a jump-diffusion equation and its numerical approximation. *SIAM Journal on Scientific Computing*, 31(2):1141–1155, 2009.
- [61] YongHoon Kwon and Younhee Lee. A second-order finite difference method for option pricing under jump-diffusion models. *SIAM journal on numerical analysis*, 49(6):2598–2617, 2011.
- [62] Maya Briani, Roberto Natalini, and Giovanni Russo. Implicit–explicit numerical schemes for jump–diffusion processes. *Calcolo*, 44(1):33–57, 2007.
- [63] Serguei Y Novak. Extreme value methods with applications to finance. *Monographs on Statistics and Applied Probability*, 122:22, 2011.
- [64] Peter G Fennell, Sergey Melnik, and James P Gleeson. Limitations of discrete-time approaches to continuous-time contagion dynamics. *Physical Review E*, 94(5):052125, 2016.
- [65] Daniel T Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1-3):404–425, 1992.
- [66] Thomas G Kurtz. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8(2):344–356, 1971.
- [67] Ilkka Hanski. Metapopulation dynamics. *Nature*, 396(6706):41–49, 1998.

- [68] Thomas G Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*, 6(3):223–240, 1978.
- [69] Marco Sarich. *Projected Transfer Operators: Discretization of Markov processes in high-dimensional state spaces*. PhD thesis, Freie Universität Berlin, 2011.
- [70] Véronique Gayraud, Anton Bovier, Michael Eckhoff, and Markus Klein. Metastability in reversible diffusion processes i: Sharp asymptotics for capacities and exit times. *Journal of the European Mathematical Society*, 6(4):399–424, 2004.
- [71] Christof Schütte and Wilhelm Huisinga. Biomolecular conformations can be identified as metastable sets of molecular dynamics. *Handbook of numerical analysis*, 10:699–744, 2003.
- [72] S Rüdrieh, Marco Sarich, and Christof Schütte. Utilizing hitting times for finding metastable sets in non-reversible markov chains. *Journal of Comp. Dynamics*, 2017.
- [73] Frank Noé, Illia Horenko, Christof Schütte, and Jeremy C Smith. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *The Journal of chemical physics*, 126(15):04B617, 2007.
- [74] Nataša Djurdjevac, Marco Sarich, and Christof Schütte. Estimating the eigenvalue error of Markov state models. *Multiscale Modeling & Simulation*, 10(1):61–81, 2012.
- [75] Christof Schütte, Frank Noé, Jianfeng Lu, Marco Sarich, and Eric Vanden-Eijnden. Markov state models based on milestoneing. *The Journal of chemical physics*, 134(20):05B609, 2011.
- [76] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J. Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.
- [77] Stefanie Winkelmann and Christof Schütte. The spatiotemporal master equation: Approximation of reaction-diffusion dynamics via Markov state modeling. *The Journal of Chemical Physics*, 145(21):214107, 2016.
- [78] Marco Sarich, Frank Noé, and Christof Schütte. On the approximation quality of Markov state models. *Multiscale Modeling & Simulation*, 8(4):1154–1177, 2010.
- [79] Stefanie Winkelmann and Christof Schütte. *Stochastic Dynamics in Computational Biology*, volume 8 of *Frontiers in Applied Dynamical Systems*. Springer, 2020.
- [80] Aurélien Alfonsi, Eric Cances, Gabriel Turinici, Barbara Di Ventura, and Wilhelm Huisinga. Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems. In *ESAIM: proceedings*, volume 14, pages 1–13. EDP Sciences, 2005.

- [81] Pierre Montagnon. Stability of piecewise deterministic Markovian metapopulation processes on networks. *Stochastic Processes and their Applications*, 130(3):1515 – 1544, 2020.
- [82] Stephan Menz. *Hybrid stochastic-deterministic approaches for simulation and analysis of biochemical reaction networks*. PhD thesis, Freie Universität Berlin, 2013.
- [83] Jon Louis Bentley and Jerome H Friedman. Data structures for range searching. *ACM Computing Surveys (CSUR)*, 11(4):397–409, 1979.
- [84] Hemant M Kakde. Range searching using kd tree. *Florida State University*, 2005.
- [85] Donald L DeAngelis and Wolf M Mooij. Individual-based modeling of ecological and evolutionary processes. *Annual Review of Ecology, Evolution, and Systematics*, pages 147–168, 2005.
- [86] Cornelia Becker, Norbert Benecke, Ana Grabundžija, Hans-Christian Küchelmann, Susan Pollock, Wolfram Schier, Chiara Schoch, Ingo Schrakamp, Britta Schütt, and Martin Schumacher. The textile revolution. research into the origin and spread of wool production between the Near East and Central Europe. In *Space and Knowledge. Topoi Research Group Articles*, volume 6, pages 102–151, 2016.
- [87] Joy McCorrison. Textile extensification, alienation, and social stratification in Ancient Mesopotamia. *Current Anthropology*, 38(4):517–535, 1997.
- [88] Mariëka Brouwer Burg, Hans Peeters, and William A Lovis. *Uncertainty and sensitivity analysis in archaeological computational modeling*. 2016.
- [89] Simon Timberlake. The use of experimental archaeology/archaeometallurgy for the understanding and reconstruction of early bronze age mining and smelting technologies. *Metals and mines: Studies in archaeometallurgy*, pages 27–36, 2007.
- [90] Metin I Eren, Michelle R Bebbler, James D Norris, Alyssa Perrone, Ashley Rutkoski, Michael Wilson, and Mary Ann Raghanti. Experimental replication shows knives manufactured from frozen human feces do not work. *Journal of Archaeological Science: Reports*, 27:102002, 2019.
- [91] Mario V Tomasello, Giacomo Vaccario, and Frank Schweitzer. Data-driven modeling of collaboration networks: a cross-domain analysis. *EPJ Data Science*, 6(1):22, 2017.
- [92] C Tineke Rooijackers. Spinning animal fibres at Late Neolithic Tell Sabi Abyad, Syria? *Paléorient*, pages 93–109, 2012.
- [93] MJ Plank and A James. Optimal foraging: Lévy pattern or process? *Journal of The Royal Society Interface*, 5(26):1077–1086, 2008.

- [94] Philip M Morse. Diatomic molecules according to the wave mechanics. ii. vibrational levels. *Physical Review*, 34(1):57, 1929.
- [95] José A Carrillo, Stephan Martin, and Vladislav Panferov. A new interaction potential for swarming models. *Physica D: Nonlinear Phenomena*, 260:112–126, 2013.
- [96] Andy Jarvis, Hannes Isaak Reuter, Andrew Nelson, Edward Guevara, et al. Hole-filled srtm for the globe version 4. *available from the CGIAR-CSI SRTM 90m Database (<http://srtm.csi.cgiar.org>)*, 2008.
- [97] PFBJ Quinn, K Beven, Pierre Chevallier, and Olivier Planchon. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological processes*, 5(1):59–79, 1991.
- [98] Jarosław Jasiewicz and Tomasz F Stepinski. Geomorphons—a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, 182:147–156, 2013.
- [99] Tomasz F Stepinski and Jaroslaw Jasiewicz. Geomorphons—a new approach to classification of landforms. *Proceedings of Geomorphometry 2011*, pages 109–112, 2011.
- [100] Guenther Fischer, Freddy Nachtergaele, Sylvia Prieler, HT Van Velthuizen, Luc Verelst, and David Wiberg. Global agro-ecological zones assessment for agriculture (gaez 2008). *IIASA, Laxenburg, Austria and FAO, Rome, Italy*, 10, 2008.
- [101] Carsten Lemmen. Mechanisms shaping the transition to farming in Europe and the North American Woodland. *Archaeology, Ethnology and Anthropology of Eurasia*, 41(3):48–58, 2013.
- [102] Ana Grabundžija and Emmanuele Russo. Tools tell tales-climate trends changing threads in the prehistoric Pannonian Plain. *Documenta Praehistorica*, 43:301, 2016.
- [103] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.
- [104] Jan-Hendrik Niemann, Stefan Klus, and Christof Schütte. Data-driven model reduction of agent-based systems using the koopman generator. *PloS one*, 16(5):e0250970, 2021.
- [105] Han Cheng Lie, Konstantin Fackeldey, and Marcus Weber. A square root approximation of transition rates for a markov state model. *SIAM Journal on Matrix Analysis and Applications*, 34(2):738–756, 2013.
- [106] J MacQueen. Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, pages 281–297, 1967.

- [107] Leland McInnes, John Healy, and Steve Astels. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.*, 2(11):205, 2017.
- [108] Michael Kamp, Linara Adilova, Joachim Sicking, Fabian Hüger, Peter Schlicht, Tim Wirtz, and Stefan Wrobel. Efficient decentralized deep learning by dynamic model averaging. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 393–409. Springer, 2018.
- [109] Daniel McNeish. On using bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal*, 23(5):750–773, 2016.
- [110] Rens van de Schoot, Sarah Depaoli, Ruth King, Bianca Kramer, Kaspar Märtens, Mahlet G Tadesse, Marina Vannucci, Andrew Gelman, Duco Veen, Joukje Willemssen, et al. Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1):1–26, 2021.
- [111] 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.
- [112] Serge Bahuchet. Changing language, remaining pygmy. *Human Biology*, 84(1):11–43, 2012.
- [113] Richard G Klein. *The human career: Human biological and cultural origins*. University of Chicago Press, 2009.
- [114] Lewis R Binford. Willow smoke and dogs’ tails: hunter-gatherer settlement systems and archaeological site formation. *American antiquity*, 45(1):4–20, 1980.
- [115] Robert L Kelly. Hunter-gatherer mobility strategies. *Journal of anthropological research*, 39(3):277–306, 1983.
- [116] Frank W Marlowe. Hunter-gatherers and human evolution. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, 14(2):54–67, 2005.
- [117] Matt Grove. Hunter-gatherers adjust mobility to maintain contact under climatic variation. *Journal of Archaeological Science: Reports*, 19:588–595, 2018.
- [118] Michael Shott. Technological organization and settlement mobility: an ethnographic examination. *Journal of Anthropological Research*, 42(1):15–51, 1986.
- [119] Mark Collard, Michael Kemery, and Samantha Banks. Causes of toolkit variation among hunter-gatherers: a test of four competing hypotheses. *Canadian Journal of Archaeology/Journal Canadien d’Archéologie*, pages 1–19, 2005.

- [120] Mark Collard, Briggs Buchanan, and Michael J O'Brien. Population size as an explanation for patterns in the paleolithic archaeological record: more caution is needed. *Current Anthropology*, 54(S8):S388–S396, 2013.
- [121] Robert Boyd, Peter J Richerson, et al. Why culture is common, but cultural evolution is rare. In *Proceedings-British Academy*, volume 88, pages 77–94. Oxford University Press Inc., 1996.
- [122] Albano Beja-Pereira, Gordon Luikart, Phillip R England, Daniel G Bradley, Oliver C Jann, Giorgio Bertorelle, Andrew T Chamberlain, Telmo P Nunes, Stoitcho Metodiev, Nuno Ferrand, et al. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nature genetics*, 35(4):311–313, 2003.
- [123] Joseph Henrich. *The secret of our success*. Princeton University Press, 2015.
- [124] Lukas S Premo. Effective population size and the effects of demography on cultural diversity and technological complexity. *American Antiquity*, 81(4):605–622, 2016.
- [125] Maxime Derex and Robert Boyd. Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, 113(11):2982–2987, 2016.
- [126] Andrea B Migliano, Federico Battiston, Sylvain Viguier, Abigail E Page, Mark Dyble, Rodolph Schlaepfer, Daniel Smith, Leonora Astete, Marilyn Ngales, Jesus Gomez-Gardenes, et al. Hunter-gatherer multilevel sociality accelerates cumulative cultural evolution. *Science advances*, 6(9):eaax5913, 2020.
- [127] Ketika Garg, Cecilia Padilla-Iglesias, Nicolás Restrepo Ochoa, and V Bleu Knight. Hunter-gatherer foraging networks promote information transmission. *Royal Society open science*, 8(12):211324, 2021.
- [128] Luigi Luca Cavalli-Sforza and Marcus W Feldman. *Cultural transmission and evolution: A quantitative approach*. Princeton University Press, 1981.
- [129] Robert Boyd and Peter J Richerson. Why does culture increase human adaptability? *Ethology and sociobiology*, 16(2):125–143, 1995.
- [130] Alex Mesoudi and Andrew Whiten. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509):3489–3501, 2008.
- [131] Marcus J Hamilton, Briggs Buchanan, and Robert S Walker. Scaling the size, structure, and dynamics of residentially mobile hunter-gatherer camps. *American Antiquity*, 83(4):701–720, 2018.
- [132] Vivek V Venkataraman, Thomas S Kraft, Nathaniel J Dominy, and Kirk M Endicott. Hunter-gatherer residential mobility and the marginal value of rainforest patches. *Proceedings of the National academy of Sciences*, 114(12):3097–3102, 2017.

- [133] Mark Collard, Briggs Buchanan, Michael J O'Brien, and Jonathan Scholnick. Risk, mobility or population size? drivers of technological richness among contact-period western north american hunter-gatherers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1630):20120412, 2013.
- [134] Elizabeth M Gallagher, Stephen J Shennan, and Mark G Thomas. Transition to farming more likely for small, conservative groups with property rights, but increased productivity is not essential. *Proceedings of the National Academy of Sciences*, 112(46):14218–14223, 2015.
- [135] Robert L Kelly. *The lifeways of hunter-gatherers: the foraging spectrum*. Cambridge University Press, 2013.
- [136] Daniel Smith, Mark Dyble, James Thompson, Katie Major, Abigail E Page, Nikhil Chaudhary, Gul Deniz Salali, Lucio Vinicius, Andrea Bamberg Migliano, and Ruth Mace. Camp stability predicts patterns of hunter-gatherer cooperation. *Royal Society Open Science*, 3(7):160131, 2016.
- [137] Mark Dyble, James Thompson, Daniel Smith, Gul Deniz Salali, Nikhil Chaudhary, Abigail E Page, Lucio Vinicius, Ruth Mace, and Andrea Bamberg Migliano. Networks of food sharing reveal the functional significance of multilevel sociality in two hunter-gatherer groups. *Current Biology*, 26(15):2017–2021, 2016.
- [138] Maxime Derex and Alex Mesoudi. Cumulative cultural evolution within evolving population structures. *Trends in Cognitive Sciences*, 24(8):654–667, 2020.
- [139] Matthew A Peeples. Finding a place for networks in archaeology. *Journal of Archaeological Research*, 27(4):451–499, 2019.
- [140] Sara Graça Da Silva and Jamshid J Tehrani. Comparative phylogenetic analyses uncover the ancient roots of indo-european folktales. *Royal Society open science*, 3(1):150645, 2016.
- [141] Robert Axelrod. The dissemination of culture: A model with local convergence and global polarization. *Journal of conflict resolution*, 41(2):203–226, 1997.
- [142] Mirna Kovacevic, Stephen Shennan, Marian Vanhaeren, Francesco d'Errico, and Mark G Thomas. Simulating geographical variation in material culture: were early modern humans in europe ethnically structured? In *Learning strategies and cultural evolution during the Palaeolithic*, pages 103–120. Springer, 2015.
- [143] Oren Kolodny, Nicole Creanza, and Marcus W Feldman. Evolution in leaps: the punctuated accumulation and loss of cultural innovations. *Proceedings of the National Academy of Sciences*, 112(49):E6762–E6769, 2015.
- [144] Jesús Olivero, Julia E Fa, Miguel A Farfán, Jerome Lewis, Barry Hewlett, Thomas Breuer, Giuseppe M Carpaneto, María Fernández, Francesco Germi, Shiho Hattori, et al. Distribution and numbers of pygmies in central african forests. *PloS one*, 11(1):e0144499, 2016.

- [145] Deborah S Rogers, Marcus W Feldman, and Paul R Ehrlich. Inferring population histories using cultural data. *Proceedings of the Royal Society B: Biological Sciences*, 276(1674):3835–3843, 2009.
- [146] W Brian Arthur. *The nature of technology: What it is and how it evolves*. Simon and Schuster, 2009.
- [147] Kevin N Laland. Social learning strategies. *Animal Learning & Behavior*, 32(1):4–14, 2004.
- [148] Zachary H Garfield, Melissa J Garfield, and Barry S Hewlett. A cross-cultural analysis of hunter-gatherer social learning. In *Social learning and innovation in contemporary hunter-gatherers*, pages 19–34. Springer, 2016.
- [149] Robert M Beyer, Mario Krapp, and Andrea Manica. High-resolution terrestrial climate, bioclimate and vegetation for the last 120,000 years. *Scientific data*, 7(1):1–9, 2020.
- [150] Publications of the US Geological Survey. Technical report, 1996.
- [151] Robert L Bettinger. Prehistoric hunter-gatherer population growth rates rival those of agriculturalists. *Proceedings of the National Academy of Sciences*, 113(4):812–814, 2016.
- [152] Derek Greene, Donal Doyle, and Padraig Cunningham. Tracking the evolution of communities in dynamic social networks. In *2010 international conference on advances in social networks analysis and mining*, pages 176–183. IEEE, 2010.
- [153] Konstantin Fackeldey, Péter Koltai, Peter Névir, Henning Rust, Axel Schild, and Marcus Weber. From metastable to coherent sets—time-discretization schemes. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(1):012101, 2019.
- [154] Péter Koltai, Giovanni Ciccotti, and Christof Schütte. On metastability and markov state models for non-stationary molecular dynamics. *The Journal of chemical physics*, 145(17):174103, 2016.
- [155] Jan-Hendrik Niemann, Stefanie Winkelmann, Sarah Wolf, and Christof Schütte. Agent-based modeling: Population limits and large timescales. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(3), 2021.
- [156] Daniel Bernoulli. Essai d’une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l’inoculation pour la prévenir. *Histoire de l’Acad., Roy. Sci.(Paris) avec Mem*, pages 1–45, 1760.
- [157] Ian Cooper, Argha Mondal, and Chris G Antonopoulos. A sir model assumption for the spread of covid-19 in different communities. *Chaos, Solitons & Fractals*, 139:110057, 2020.

- [158] Hanna Wulkow, Tim Conrad, Nataša Djurdjevac Conrad, Sebastian A. Mueller, Kai Nagel, and Christof Schuette. Prediction of covid-19 spreading and optimal coordination of counter-measures: From microscopic to macroscopic models to pareto fronts. *medRxiv*, 2020.
- [159] Ling Xue, Shuanglin Jing, Joel C Miller, Wei Sun, Huafeng Li, José Guillermo Estrada-Franco, James M Hyman, and Huaiping Zhu. A data-driven network model for the emerging covid-19 epidemics in wuhan, toronto and italy. *Mathematical biosciences*, 326:108391, 2020.
- [160] Daniela Calvetti, Alexander P Hoover, Johnie Rose, and Erkki Somersalo. Metapopulation network models for understanding, predicting, and managing the coronavirus disease covid-19. *Frontiers in Physics*, 8:261, 2020.
- [161] Sebastian A Müller, Michael Balmer, William Charlton, Ricardo Ewert, Andreas Neumann, Christian Rakow, Tilmann Schlenther, and Kai Nagel. Predicting the effects of covid-19 related interventions in urban settings by combining activity-based modelling, agent-based simulation, and mobile phone data. *PloS one*, 16(10):e0259037, 2021.
- [162] Francesco Messina, Emanuela Giombini, Chiara Agrati, Francesco Vairo, Tommaso Ascoli Bartoli, Samir Al Moghazi, Mauro Piacentini, Franco Locatelli, Gary Kobinger, Markus Maeurer, et al. Covid-19: viral–host interactome analyzed by network based–approach model to study pathogenesis of sars-cov-2 infection. *Journal of translational medicine*, 18(1):1–10, 2020.
- [163] Jos Lelieveld, Frank Helleis, Stephan Borrmann, Yafang Cheng, Frank Drewnick, Gerald Haug, Thomas Klimach, Jean Sciare, Hang Su, and Ulrich Pöschl. Model calculations of aerosol transmission and infection risk of covid-19 in indoor environments. *International Journal of Environmental Research and Public Health*, 17(21):8114, 2020.
- [164] Dominik Ziemke, Ihab Kaddoura, and Kai Nagel. The matsim open berlin scenario: A multimodal agent-based transport simulation scenario based on synthetic demand modeling and open data. *Procedia computer science*, 151:870–877, 2019.
- [165] A Das, S Ghosh, K Das, T Basu, M Das, and I Dutta. Modeling the effect of area deprivation on covid-19 incidences: a study of chennai megacity, india. *Public Health*, 185:266–269, 2020.
- [166] Morteza Abdullatif Khafaie and Fakher Rahim. Cross-country comparison of case fatality rates of covid-19/sars-cov-2. *Osong Public Health and Research Perspectives*, 11(2):74, 2020.
- [167] Richard Levins. Some demographic and genetic consequences of environmental heterogeneity for biological control. *American Entomologist*, 15(3):237–240, 1969.

- [168] Wan Yang, Sasikiran Kandula, Mary Huynh, Sharon K Greene, Gretchen Van Wye, Wenhui Li, Hiu Tai Chan, Emily McGibbon, Alice Yeung, Don Olson, et al. Estimating the infection-fatality risk of sars-cov-2 in new york city during the spring 2020 pandemic wave: a model-based analysis. *The Lancet Infectious Diseases*, 21(2):203–212, 2021.
- [169] Frank Schlosser, Benjamin F. Maier, Olivia Jack, David Hinrichs, Adrian Zachariae, and Dirk Brockmann. Covid-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proceedings of the National Academy of Sciences*, 117(52):32883–32890, 2020.
- [170] Joshua S Weitz and Jonathan Dushoff. Modeling post-death transmission of ebola: challenges for inference and opportunities for control. *Scientific reports*, 5:8751, 2015.
- [171] Anca Rădulescu, Cassandra Williams, and Kieran Cavanagh. Management strategies in a seir-type model of covid 19 community spread. *Scientific Reports*, 10(1):21256, Dec 2020.
- [172] Sahamoddin Khailaie, Tanmay Mitra, Arnab Bandyopadhyay, Marta Schips, Pietro Mascheroni, Patrizio Vanella, Berit Lange, Sebastian C. Binder, and Michael Meyer-Hermann. Development of the reproduction number from coronavirus sars-cov-2 case data in germany and implications for political measures. *BMC Medicine*, 19(1):32, Jan 2021.
- [173] Hyeontae Jo, Hwijae Son, Se Young Jung, and Hyung Ju Hwang. Analysis of covid-19 spread in south korea using the sir model with time-dependent parameters and deep learning. *medRxiv*, 2020.
- [174] Kate M Bubar, Kyle Reinholt, Stephen M Kissler, Marc Lipsitch, Sarah Cobey, Yonatan H Grad, and Daniel B Larremore. Model-informed covid-19 vaccine prioritization strategies by age and serostatus. *Science*, 371(6532):916–921, 2021.
- [175] Matt J Keeling, Edward M Hill, Erin E Gorsich, Bridget Penman, Glen Guyver-Fletcher, Alex Holmes, Trystan Leng, Hector McKimm, Massimiliano Tamborino, Louise Dyson, et al. Predictions of covid-19 dynamics in the uk: short-term forecasting and analysis of potential exit strategies. *PLoS computational biology*, 17(1):e1008619, 2021.
- [176] Andrew W Byrne, David McEvoy, Aine Collins, Kevin Hunt, Miriam Casey, Ann Barber, Francis Butler, John Griffin, Elizabeth Lane, Conor McAloon, et al. Inferred duration of infectious period of SARS-CoV-2: rapid scoping review and analysis of available evidence for asymptomatic and symptomatic COVID-19 cases. *medRxiv*, 2020.
- [177] Weier Wang, Jianming Tang, and Fangqiang Wei. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in wuhan, china. *Journal of medical virology*, 92(4):441–447, 2020.

- [178] Natalie M Linton, Tetsuro Kobayashi, Yichi Yang, Katsuma Hayashi, Andrei R Akhmetzhanov, Sung-mok Jung, Baoyin Yuan, Ryo Kinoshita, and Hiroshi Nishiura. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of clinical medicine*, 9(2):538, 2020. <https://doi.org/10.3390/jcm9020538>.
- [179] Robert Koch Institute. *Robert Koch Institute: SARS-CoV-2 Steckbrief zur Coronavirus-Krankheit-2019 (COVID-19) Stand: 29.5.2020*, 2020 (accessed June 10, 2020). [https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\\_Coronavirus/Steckbrief.html?nn=13490888](https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html?nn=13490888).
- [180] Worldometers.info. *Worldometers.info: Coronavirus (COVID-19) Mortality Rate, 2020* (accessed June 10, 2020). <https://www.worldometers.info/coronavirus/coronavirus-death-rate/>.
- [181] Jesús Fernández-Villaverde and Charles I Jones. Estimating and simulating a SIRD model of COVID-19 for many countries, states, and cities. Technical report, National Bureau of Economic Research, 2020.
- [182] Yousef Alimohamadi, Maryam Taghdir, and Mojtaba Sepandi. The estimate of the basic reproduction number for novel coronavirus disease (COVID-19): a systematic review and meta-analysis. *Journal of Preventive Medicine and Public Health*, 2020.
- [183] Rujin Huang, Yunju Nie, Linghua Duo, Xiaoping Zhang, Zhenhua Wu, and Jiacheng Xiong. Construction land suitability assessment in rapid urbanizing cities for promoting the implementation of united nations sustainable development goals: a case study of nanchang, china. *Environmental Science and Pollution Research*, 28(20):25650–25663, 2021.
- [184] Kay W Axhausen, Andreas Horni, and Kai Nagel. *The multi-agent transport simulation MATSim*. Ubiquity Press, 2016.
- [185] Bjoern Goldenbogen, Stephan O Adler, Oliver Bodeit, Judith AH Wodke, Aviv Korman, Lasse Bonn, X Escalera-Fanjul, Johanna EL Haffner, M Karnetzki, M Krantz, et al. Geospatial precision simulations of community confined human interactions during sars-cov-2 transmission reveals bimodal intervention outcomes. *medRxiv*, 2020.
- [186] Frank Ball, Tom Britton, Thomas House, Valerie Isham, Denis Mollison, Lorenzo Pellis, and Gianpaolo Scalia Tomba. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63–67, 2015.
- [187] Nazri Che Dom, Zulkiflee Abd Latif, Abu Hassan Ahmad, Rodziah Ismail, and Biswajeet Pradhan. Manifestation of gis tools for spatial pattern distribution analysis of dengue fever epidemic in the city of subang jaya, malaysia. *EnvironmentAsia*, 5(2), 2012.

- [188] Robert Moss, Elham Naghizade, Martin Tomko, and Nicholas Geard. What can urban mobility data reveal about the spatial distribution of infection in a single city? *BMC public health*, 19(1):1–16, 2019.
- [189] Nico Stollenwerk, Sander van Noort, José Martins, Maíra Aguiar, Frank Hilker, Alberto Pinto, and Gabriela Gomes. A spatially stochastic epidemic model with partial immunization shows in mean field approximation the reinfection threshold. *Journal of Biological Dynamics*, 4(6):634–649, 2010.
- [190] Aditya Firman Ihsan. Data-driven identification of compartmental model of covid-19. In *2021 International Conference on Data Science and Its Applications (ICoDSA)*, pages 91–96. IEEE, 2021.