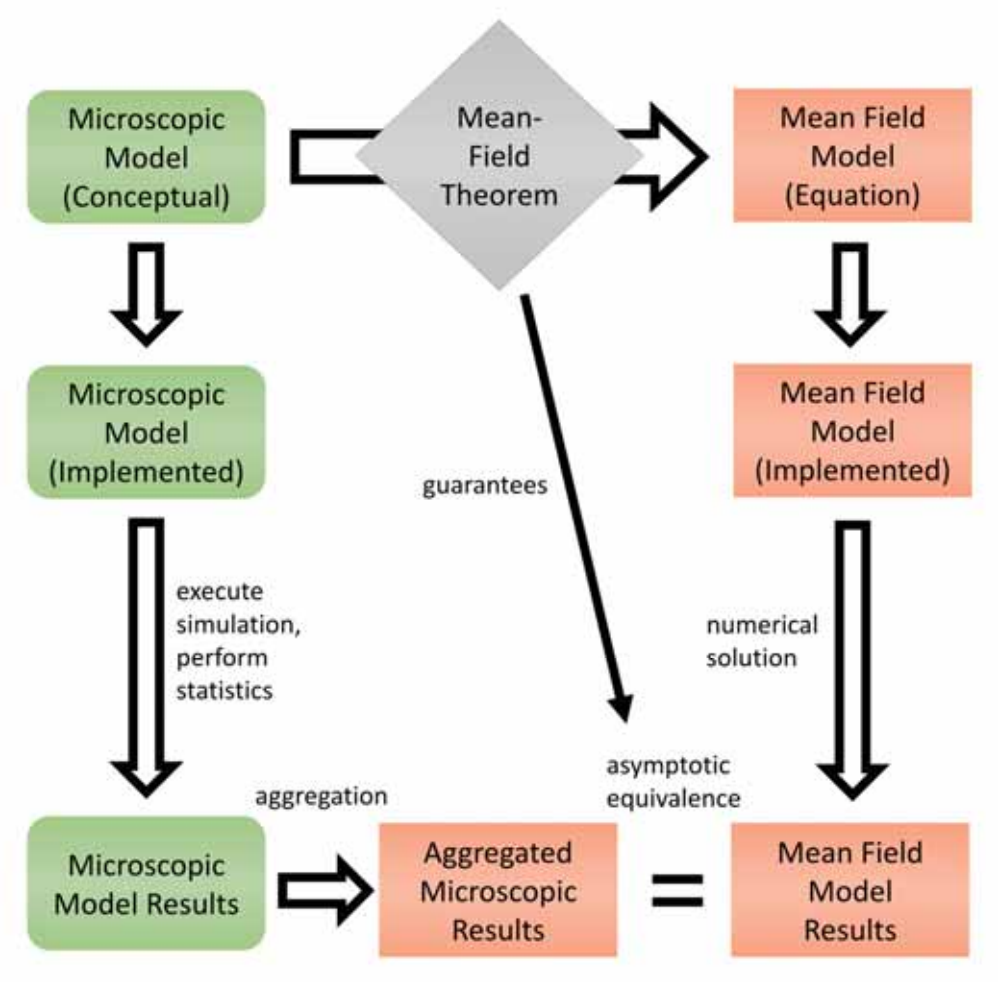
FBS 29

Fortschrittsberichte Simulation
Advances in Simulation

Classification of Microscopic Models with Respect to Aggregated System Behaviour

Martin Bicher



ISBN ebook 978-3-903347-29-8 DOI 10.11128/fbs.29

TU Verlag ARGESIM


sim


ARGESIM

FBS Fortschrittsberichte Simulation
Advances in Simulation

Herausgegeben von **ASIM - A**rbeitsgemeinschaft **Sim**ulation, Fachausschuss der **GI** – Gesellschaft für Informatik - im Fachbereich **ILW** – Informatik in den Lebenswissenschaften
Published by **ASIM** – German Simulation Society, Section of **GI** – German Society for Informatics - in Division **ILW** – Informatics in Life Sciences

ASIM FBS 29

Martin Bicher

Classification of Microscopic Models with Respect to Aggregated System Behaviour

FBS - Fortschrittsberichte Simulation / Advances in Simulation

Herausgegeben von **ASIM** - Arbeitsgemeinschaft Simulation, → *www.asim-gi.org*

ASIM ist ein Fachausschuss der **GI** - Gesellschaft für Informatik, → *www.gi.de*, im Fachbereich **ILW** – Informatik in den Lebenswissenschaften, → *fb-ilw.gi.de*

Published on behalf of **ASIM** – German Simulation Society, → *www.asim-gi.org*

ASIM is a section of of **GI** – German Society for Informatics, → *www.gi.de*, in division **ILW** – Informatics in the Life Sciences, → *fb-ilw.gi.de*

Reihenherausgeber / Series Editors

Prof. Dr.-Ing. Th. Pawletta (ASIM), HS Wismar, Thorsten.pawletta@hs-wismar.de
Prof. Dr. D. Murray-Smith (EUROSIM / ASIM), Univ. Glasgow,
David.Murray-Smith@glasgow.ac.uk
Prof. Dr. F. Breitenecker (ARGESIM / ASIM), TU Wien, Felix.Breitenecker@tuwien.ac.at

Titel / Title: Classification of Microscopic Models with Respect to Aggregated System Behaviour

Autor / Author: Martin Bicher, *ORCID ID 0000-0002-1362-6868*
Martin.Bicher@dwh.at, Martin.Bicher@tuwien.ac.at

FBS Band / Volume: 29

Typ / Type: Dissertation

ISBN ebook: 978-3-903347-29-8, ARGESIM Publisher Vienna, 2020;
www.argesim.org

DOI: 10.11128/fbs.29

Seiten / Pages: 266

Front Cover Grafics: Figure 4.2 of Thesis
Back Cover Picture of Author: ©Stefan Knittel /dwh GmbH

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna | Austria

DISSERTATION

Classification of Microscopic Models with Respect to Aggregated System Behaviour

Ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Ao.Univ.Prof.i.R. Dipl.-Ing. Dr.techn. Felix Breitenecker
101 Institut für Analysis und Scientific Computing

und

Dr.techn. Nikolas Popper
101 Institut für Analysis und Scientific Computing

eingereicht an der

Technischen Universiät Wien
Fakultär für Mathematik und Geoinformation

von

Dipl.-Ing. Martin Bicher, BSc.
Matr.-Nr. 00625429
Pfeilgasse 3/38, 1080 Vienna, Austria.

Wien, am

Unterschrift des Authors

Kurzfassung

Auf Grund der ständig stiegenden Datenmenge in unterschiedlichsten Wissenschaftsdisziplinen haben datengetriebene Methoden wie Machine Learning oder Regressionsanalysen einen hohen Stellenwert erreicht. Daher werden an mikroskopische mikroskopische Simulationsmodelle enorme Anforderungen gestellt. Aktuell finden sich solche Modelle zumeist als Nischenlösung für Fälle, wo die Datenlage schwierig und die Aufgabenstellungen zu komplex sind um sie mit datengetriebenen Methoden zu erfassen. Leider ist auch diese Position mittlerweile stark gefährdet: Einerseits erwarten sich Entscheidungsträger quantitative Antworten auf komplexe Forschungsfragen trotz geringer Verfübarkeit von Validierungsdaten, andererseits erschweren altbekannte Probleme mit Modellvalidierung, Kalibrierung, Reproduzierbarkeit und Sensitivität den Einsatz von mikroskopischen Modellen.

Aus diesem Grund ist diese Arbeit der Weiterentwicklung von mathematischen Analysemethoden von mikroskopischen Modellen gewidmet, über die Kapitel 2 einen detaillierten Überblick gibt. Nur mithilfe formaler Analyse eines Modells ist man in der Lage, dessen Systemverhalten in seiner Gesamtheit zu verstehen und damit die angesprochene Probleme der mikroskopischen Modellbildung zu lösen. Entscheidungsunterstützungs-Tool argumentieren.

Zu diesem Zweck werden Methoden aus dem Bereich der Mean-Field Theory (Molekularfeldanalyse) verwendet. Dieser, aus dem Bereich der Statistischen Mechanik stammende Wissenschaftszweig fußt auf der Idee, die aggregierten Größen eines Modells mit interagierenden subatomaren Partikeln mithilfe eines makroskopischen Modells, üblicherweise eines Differentialgleichungsmodells, zu approximieren. Mithilfe dieses Mean-Field Modells lassen sich nun bewährte Analysemethoden von gleichungsbasierten Modellen verwenden, die Prozesse im mikroskopischen Modell zu untersuchen.

Der Mehrwert der Mean-Field Theorie für angesprochene Probleme der mikroskopischen Modellierung in den Sozialwissenschaften liegt nun darin, dass ihr Einsatzbereich nicht auf die Statistische Physik eingeschränkt, sondern mit gewissen Modifikationen auf viele andere Anwendungsbereiche erweiterbar ist. Aus diesem Grund beschäftigt sich ein großer Teil dieser Arbeit mit speziell für diese Anwendungsfälle adaptierten Mean-Field Theoremen. Viele dieser Methoden, insbesondere Theoreme 4.1 und 5.1 mit Beweisen im Anhang und die Benutzerfreundlichen Step-by-Step Anleitungen in Abschnitt 4.6 sind eigene Beiträge zu diesem Wissenschaftszweig.

Im weiteren wird anhand einer Selektion von Test Cases illustriert wie diese zur Analyse von mikroskopischen Modellen verwendet werden können und wie sich direkter Added Value für Modellkalibrierung (5.6.3) und Sensitivitätsanalyse (5.4.3) generieren lässt. Die Cases erstrecken sich über unterschiedlichste Anwendungsbereiche wie Populationsdynamik (5.6.1,5.2.3), Epidemiologie (5.5.2,5.2.1), sowie Spieltheorie (5.4.1) oder Physik (5.5.1).

Motiviert von den Ergebnissen der Mean-Field Analysen wird abschließend argumentiert, dass die aktuell gängigen Fachtermini zur Beschreibung mikroskopischer Modelle nicht unbedingt für eine System-theoretisch orientierte Kommunikation der Modelle geeignet ist. Aus diesem Grund wird in Kapitel 6 eine alternative Klassifikation vorgeschlagen, die auf die Vermittlung von Modellverhaltens-relevanter Information abzielt. Diese stellt zudem das Hauptresultat der Arbeit dar.

Summary

Microscopic simulation models in social sciences nowadays have to meet high demands if they want to compete with data-based approaches like machine learning. The main reason for that is the rich availability of data for almost all human-centred research problems. Unfortunately, features like reproducibility, validation, verification, calibration and sensitivity of microscopic simulation models raise problems, which make these approaches hardly applicable for today's quantitative research problems. Consequently, we dedicate this work to the improvement of methods for mathematical analysis of microscopic simulation models. Only by thoroughly investigating microscopic models on a formal base we are able to fully understand and characterise their behaviour and finally find reasonable answers to mentioned problems. A detailed summary of microscopic modelling approaches is given in Chapter 2.

We use so-called mean-field analysis for investigating the aggregated numbers of these models. Originating from statistical physics, this method is basically used to describe the aggregated behaviour of physical models with a large number of interacting physical sub-molecular particles. Hereby, the aggregated numbers of the model are approximated by a simpler macroscopic model, a mean-field model, usually an ordinary or partial differential equation. Most importantly for this work, the concept can partially be used to describe the aggregated numbers in microscopic simulation models for completely different approaches as well. Hereby, interacting individuals like persons, cars or animals take the place of the particles.

Obviously, the usage of mean-field theory for social science applications requires slightly modified methods, as physical particles behave different than entities in microscopic models in social science. Therefore we present a couple of mean-field theorems, specifically developed for these applications, in this work. Especially by the novel statement and proof of Theorems 4.1 and 5.1 we contribute to the state of the art in complexity theory.

Using a valid mean-field model, i.e. a macroscopic equation-based model that approximates the microscopic model results, we utilize the excellent analysis features of equation based approaches to analyse the microscopic model. We will use case studies from population dynamics (5.6.1, 5.2.3) epidemiology (5.5.2, 5.2.1), Physics (5.5.1) and Game Theory (5.4.1). We additionally show how mean-field models can be applied directly for sensitivity analysis (5.4.3) and parameter calibration (5.6.3) of microscopic models and give hints to other fields of application.

We finally argue that the currently used terminology to label microscopic simulation models are not optimally suited to convey a scientific picture of the model. Thus, we conclude with the main result of this thesis in Chapter 6 and propose a new classification concept for microscopic models: A series of attributive adjectives according to the model's time-update, state-space, randomness and interaction do not only convey a unique picture of specific parts of the model, but also give ideas on possible challenges involved with model, simulation, parametrisation, sensitivity, and finally its mean-field behaviour.

Acknowledgements

Da es Schätzungen zufolge im Österreichischen mehr als achtmal soviel Worte wie im Englischen gibt, existieren darin auch viel mehr Varianten, darin Dank auszudrücken. Darum habe ich mich entschlossen diese Dankesworte in meiner Muttersprache zu verfassen.

Im Zeitraum meiner Arbeiten an dieser Dissertation haben mich viele Personen einerseits unterstützt, andererseits maßgeblich beeinflusst. Die meisten, die diese Worte hier aktiv lesen gehören zu dieser erlesenen Menge, doch möchte ich auf einige noch besonders eingehen.

- Mein mit Sicherheit größter Dank gilt eindeutig meiner Freundin Julia. Zum einen, für ihr "da Sein" als beste Freundin und Partnerin, zum anderen, für ihre unablässige Hilfe bei der Fertigstellung der Arbeit. Immerhin hat sie in den letzten Wochen sogar auf ihre geliebten sportlichen Aktivitäten verzichtet um sich mit Korrekturarbeiten beschäftigen zu können. Zitat: *Se english is not se yellow of se egg*.
- Ein extra großes "Dankeschön" möchte ich auch an meine beiden Chefs, Förderern und Betreuern Felix und Niki aussprechen. Beide waren in den letzten Jahren stets darauf bedacht mir ihre Blickpunkte auf die Forschung zu vermitteln: Einerseits, der Blick durch ein Bierglas auf die wichtige Rolle der Anwendbarkeit der Forschung, andererseits der Blick durch ein Sektglas auf die Freiheiten der Grundlagenforschung. Dahingehend will ich diese Arbeit auch dem Harmonieren von Champagner und Leberkäsesemmel widmen.
- Großer Dank gebührt zudem meinen langjährigen Arbeitskolleginnen und Arbeitskollegen in der TU und in der Drahtwarenhandlung. Unter ihnen möchte ich vor allem Stefanie hervorheben, deren wahnsinniges Vorhaben der mathematischen Korretur der Arbeit es zu verdanken ist, dass so manche inkonsistente Formel jetzt korrekter ist als vorher. Auch den Mitgliedern des avantgardistischen Arbeitszirkels FAMS möchte ich besonders danken, da sie für mich die Begriffe *Dienstreise*, *Mittagspause* oder *Besprechung* völlig neu definiert haben.
- Ganz herzlich möchte ich mich auch bei meinen Eltern bedanken. Auch wenn es für sie oft nicht einfach war, haben sie mich wo es nur ging immer unterstützt und ein offenes Ohr für mich gehabt. Auch meinem verstorbenen Großvater möchte ich hiermit Dank aussprechen, der auch auf die eine oder andere Art für diese Arbeit mitverantwortlich ist. Ich freue mich schon auf das nächste *Freitagsbier*.
- Auch möchte ich meinen Freundinnen und Freunden für ihre Unterstützung danken, insbesondere Mary, Fabian, Richard und Paul, aber auch all meinen Pfadfinderkolleginnen und Kollegen. Ich freue mich darauf, jetzt wieder ein wenig mehr Zeit mit ihnen verbringen zu können.
- Zu guter Letzt möchte ich auch allen Bachelor- und Diplom- StudentInnen danken, die innerhalb der letzten Jahre unter meinen Fittichen großartige Arbeiten hervorbrachten, denn auf vielen von ihnen konnte ich in dieser Dissertation aufbauen.

This thesis was financed by DEXHELPP. K-Project DEXHELPP is supported by BMVIT, BMWFW and the state of Vienna via COMET - Competence Centers for Excellent Technologies. Programme COMET is processed by FFG.

Contents

Contents v

1 Introduction **1**
1.0.1 Role of Microscopic Modelling and Simulation 1
1.0.2 Abstract vs. Applied Model 2
1.0.3 About this Thesis 3

2 Microscopic Modelling **5**
2.1 Microscopic Modelling Methods 7
2.1.1 Monte Carlo Simulation 7
2.1.2 Cellular Automata 12
2.1.3 Agent-Based Models 15
2.1.4 Microsimulation Models 19
2.1.5 Discrete Event Models 21
2.1.6 Discrete Event System Specification (DEVS) 24
2.1.7 Concluding Comments about Microscopic Modelling Terminology . . 25
2.2 Simulation Methods for Microscopic Models 26
2.2.1 Microscopic Simulation - Hard Coded 27
2.2.2 Microscopic Simulators 29
2.3 General Properties of Microscopic Models 30
2.3.1 Emergent Behaviour 30
2.3.2 Flexibility 34
2.3.3 Natural Description of the System 37

3 Stochastic Methods **39**
3.1 Probabilistic Basics 40
3.1.1 Probability and General Concepts 40
3.1.2 Random Variables 43
3.1.3 Density Functions 44
3.1.4 Mean, Variance and Covariance 45
3.1.5 Conditional Probabilities and Bayesian Theorem 48
3.1.6 Conditional Expectation and Covariance 50
3.2 Statistical Basics 51

3.2.1 Standard Estimators for Moments . . . 52
3.2.2 Estimators for Probability- and Density- Functions . . . 53
3.3 Limit Theorems of Stochastic Sums . . . 55
3.3.1 Law of Large Numbers . . . 56
3.3.2 Law of the Iterated Logarithm . . . 57
3.3.3 Central Limit Theorem . . . 58
3.3.4 Extensions and Limitations of the Three Limit Theorems . . . 59
3.4 Stochastic Processes . . . 61
3.4.1 Classification with Respect to Space Structure . . . 62
3.4.2 Classification with Respect to Memory . . . 63
3.4.3 Transition Probabilities . . . 66
3.4.4 Transition Rates . . . 67

4 Mean Field Analysis **71**
4.1 Formalised Microscopic Model . . . 72
4.1.1 Formalised Microscopic Model for Static Population Models . . . 74
4.1.2 Formalised Microscopic Model for Dynamic Population Models . . . 76
4.2 Introduction to Mean-Field Theorems . . . 78
4.2.1 Aggregated Numbers . . . 78
4.2.2 Concepts of Mean-Field Theorems . . . 79
4.3 General Mean-Field Theorem for Static Population Models . . . 80
4.3.1 Prerequisites for MFTs for Static Population Models . . . 80
4.3.2 Statement of the MFT for Static Population Models . . . 83
4.4 General MFT for Dynamic Population Models . . . 84
4.4.1 Prerequisites for MFTs for Dynamic Population Models . . . 84
4.4.2 Statement of the MFT for Dynamic Population Models . . . 86
4.5 Mean-Field Equations for Specific State Spaces . . . 87
4.5.1 Mean-Field Equations for Discrete State Space . . . 87
4.5.2 Mean-Field Equations for Continuous State Space . . . 88
4.5.3 Mean-Field Theorems for Hybrid State Space . . . 92
4.5.4 Variance Theorem for Discrete State Space . . . 94
4.6 Usage of Mean-Field Theorems . . . 98
4.6.1 General Applicability of Mean-Field Theorems . . . 98
4.6.2 Step By Step from Microscopic to Macroscopic . . . 99
4.6.3 Scientific Background of the Micro-to-Macro Instruction . . . 114
4.6.4 Step By Step from Macroscopic to Microscopic . . . 115
4.6.5 Scientific Background of the Macro-to-Micro Instruction . . . 121

5 Application of Mean-Field Theorems – Case Studies **123**
5.1 Different Time Concepts in Microscopic Models . . . 124
5.1.1 Time Scaling of Stochastic Models . . . 124
5.1.2 Discrete to Continuous Time . . . 127
5.1.3 Scaling Case Study: Microscopic Population Model . . . 131
5.2 Case Studies for the Step-By-Step Processes . . . 133

5.2.1 Micro-to-Macro Process Case Study: SIR Lattice Model 135
5.2.2 Variance Theorem Case Study: SIR Lattice Model 140
5.2.3 Macro-to-Micro Process Case Study: Modified Levins Model 143
5.3 Mean-Field Analysis for Lattice Models 150
5.3.1 Neighbourhood Concepts in Lattice Models 150
5.3.2 Neighbourhood Case Study: Logistic Growth Lattice Model 152
5.3.3 1D-Neighbourhood Case Study: Logistic Growth Lattice Model 155
5.3.4 Pair Approximation Case Study: Logistic Growth Lattice Model 159
5.4 Mean-Field Analysis for Microscopic Models with High Sensitivity 168
5.4.1 Sensitivity Case Study: Game of Life 169
5.4.2 Failed Mean-Field Analysis Case Study: Game of Life 173
5.4.3 Sensitivity Case Study: Shuffled Game of Life 176
5.5 Mean-Field Analysis for State-Continuous Models 181
5.5.1 Case Study: Random Walk Model for Groundwater Pollution 181
5.5.2 Case Study: Age-Dependent SIR Model 186
5.5.3 Interpretation of Aggregated Numbers in Space Continuous Models . . 190
5.6 Mean-Field Analysis for Parametrised Models 194
5.6.1 Case Study: Generic Population Concept (GEPOC) 196
5.6.2 Mean-Field Analysis of GEPOC ABM 199
5.6.3 Calibration Case Study: Morbidity Extension of GEPOC ABM 205

6 Classification of Microscopic Models 211
6.1 Classification Parameters 212
6.1.1 Classification with Respect to Time-Update 212
6.1.2 Classification with Respect to State-Space 214
6.1.3 Classification with Respect to Randomness 216
6.1.4 Classification with Respect to Interaction Level 217
6.2 Classification Summary 220

7 Conclusion and Outlook 225

A Appendix 227
A.0.1 Proof for the Dynamic General Mean-Field Theorem 4.1 227
A.0.2 Proof for the Discrete-Space Variance Theorem 5.1 234

List of Figures 243

Bibliography 247

CHAPTER 1

Introduction

1.0.1 Role of Microscopic Modelling and Simulation

Since about ten years so-called data-based approaches like machine-learning, data-visualisation, data-mining or other statistical approaches undoubtedly count to the most significant methods for answering research questions. The amount of data collected nowadays all over the world is clearly enormous, mostly due to high usage of interconnected electronic devices. According to Jim Grey [Hey et al., 2009] data-based approaches supersede many classical modelling and simulation methods in usability and efficacy.

Nowadays classical computational simulation models become much less frequently used than in the course of the 1990s and 2000s as it is much more difficult to make use of collected data with them. Especially the once very successful research on microscopic simulation methods like agent-based- or microsimulation- modelling almost seems to stagnate, as these methods are rarely able to compete with data-based approaches for equivalent research questions, if sufficient amount of data is available.

Yet it is too early to postulate (like done in [Hey et al., 2009]) that computational modelling and simulation is an outdated concept, as there is still need for this method. The lack of data due to ethical reasons (e.g. medication tests) or (current) immeasurability of specific target values (e.g. forecasting of complex processes), high data bias, bad data quality (e.g. heterogeneously reported data), or data of complex interacting processes in general, to list only a few of the fields where modelling and simulation is still needed. In these cases modelling and simulation can still be used to gather additional information by adding **causal** information in form of hypotheses while data-based approaches fail.

Therefore, requirements for simulation models are immense as data- availability is bad, but processes are complex:

1. Simulation models need to be compatible with any available data that might improve the quality/validity of the result.

2. Simulation models need to be able to validly depict very complex, non-linear processes.

3. Simulation model results are expected to be quantitatively valid (as this is expected from data-based approaches as well), yet required validation data is often missing.

While 1 is a very technical challenge, the other two are related to the question: **How well do I know my model?** Only if we know precisely how causal hypotheses influence the model's behaviour, we can improve the results from data-based approaches and generate added value via causality. Hence, formal model **analysis** is crucial.

The analysis of (simulation) models with respect to their behaviour is a very old and rigorously researched field of science, which dates back to the first mathematicians dealing with differential equations. Commonly the names Isaac Newton, Gottfried Leibniz and the Bernoulli family are stated as founding fathers of this field of research. In these times the model analysis was clearly prioritised over the model execution (i.e. the simulation), for the latter was, in most cases, not possible for obvious reasons. This is also true for the first appearance of (simulation) models in human-centred sciences which have been established long before the invention of computers or other automatised calculators (e.g. the SIR model by Kermack and McKendrik in 1927 [Kermack and McKendrick, 1927]). Finally, the invention of computers and especially the great improvement of computer performance in the 1990s made it possible to quickly simulate any defined model. Therefore, formal model analysis became less important, as the models became larger and more complex and also because it was deemed to be of minor importance.

Especially the growing popularity of **microscopic modelling approaches** in that period raised issues with respect to model analysis. Due to their flexibility and their algorithmical description language they are perfectly suited nowadays to validly answer complex research problems. Nevertheless, these features also make it very difficult to formalise and, therefore, analyse them with analytical means.

It is anyway difficult to argue why a defined microscopic simulation model should be chosen in favour of a data-based approach, if the behaviour of the model, i.e. the mechanism that contains the causal information, are not understood. Both approaches are essentially black-boxes of which mostly the data-based approach is the more credible one. For this reason, we dedicate this work to contribute to methods for analysing microscopic approaches.

1.0.2 Abstract vs. Applied Model

In terms of model analysis in general it is crucial to separate the formalised model from its research question and its modelling process. We may call a model detached from its research question as an *abstract model*, while the model in the context of the real reference system will be denoted as *applied model* in this section. Information gathered for a specific abstract model clearly is reusable for any applied model that is based on the same abstract model, or, at least, on parts of it.

One of the best examples to demonstrate this is the process of heat-transfer trough a solid medium. Investigation of the real system classically leads to the following second order partial differential equation

$$\frac{\partial}{\partial t}u = D\Delta u. \tag{1.1}$$

Formal analysis of this equation is required to gain insights into the heat-transfer problem. Yet it can also be used to get an impression of the related diffusion problem wherein diffusion of liquids is investigated. Hence, also diffusion problems can be modelled validly with the diffusion-equation. Summarising, both diffusion and heat-transfer, can be modelled with the same abstract model (note that the equation above is classically called diffusion- or heat-transfer- equation). Thus, analysis of the latter can be used for both problems.

It is sometimes quite difficult to "distil" an abstract core from an applied model. For microscopic models this is especially difficult, as the models are extremely heterogeneously defined. This is emphasised by the fact that there is almost no microscopic modelling method that has a unique formal definition all modelling scientist agree with (e.g. agent-based modelling, cellular automata, etcetera. See Chapter 2).

In general, mentioned distillation process is considerably easier if models can be classified with respect to specific properties. Hence, we define the following research question for this thesis.

Definition 0.1: Research Question

In this thesis we aim to define a classification of the field of microscopic models. The classification-attributes are chosen for their specific influence on the aggregated behaviour of the model.

With this work we will contribute to improve understanding of the processes that lead to specific behaviour in microscopic models. We use and enhance methods from mean-field analysis as well as specifically chosen test-cases to reinforce the choice of classes.

1.0.3 About this Thesis

We present a broad overview of the thesis to give guidance concerning the intentions of the different chapters. Hereby the readers get an impression which chapters might be of interest for them.

- Chapter 2 gives a very broad overview on different microscopic modelling methods. This overview partly explains **why** we currently have problems with the analysis of microscopic models from a historical background. We will focus on the large number of different subclasses of microscopic modelling methods in 2.1 and present the origin of each of them. That way it will become clear why we nowadays don't have a common wording all can agree on. In Section 2.3 we focus on the challenges and features of microscopic modelling in general.

- Chapter 3 gives a very basic overview of important stochastic methods to introduce the analysis methods which will be used and enhanced later in this work. Starting with defining random variables through to the Law of the Iterated Logarithm, a very broad summary of probability theory, statistics and stochastic processes is given.The main idea of this chapter is not only to summarise a base of commonly known definitions and theorems

while explaining the more advanced mean-field methods, but also to introduce a formalism for probability concepts that is used in this thesis.

- Our contribution to the complexity-related scientific field of Mean-Field Theory, which is used for formal analysis and, finally, the classification of microscopic models, can be found in Chapter 4. Herein we will state a couple of theorems and corollaries which are proven in cited literature or in the Appendix, if they are a new contribution.

  In Section 4.6 two step-by-step instructions can be found which are defined to simplify the mathematically very demanding concept of mean-field analysis for people which are less familiar with advanced stochastic formalisms.

- Chapter 5 includes a couple of specifically chosen test cases. First, we aim to show how mean-field analysis concepts and especially the step-by-step instructions defined in the prior chapter are used to analyse the behaviour of microscopic models. We will show examples developed with different microscopic modelling methods and originated in different research fields to show the general validity of the mean-field methods.

  Secondly, we will show a couple of direct applications of mean-field analysis apart from mentioned classification. We will focus on sensitivity, calibration and interpretation of simulation output.

  Thirdly, the test cases will motivate the choice of the classification of microscopic models.

- Finally, mentioned classification of microscopic models can be found in Chapter 6. We will argue why this classification is helpful for analysis and communication of a microscopic model and give examples for how the classification can be used. This is done for a selection of well known microscopic models.

CHAPTER 2

Microscopic Modelling

The term **microscopic** is based on the Greek words *mikros* (≈ small) and *skopeo* (≈ see or look), and therefore can be translated roughly into "looking on small things". Consequently, a microscopic model could be denoted as a model that depicts small things. Clearly, **macroscopic** can be seen as of opposite of microscopic as it means to "look at large things". As "small" and "large" always depends on the observer of the system and the chosen scope we can deduct:

- The term microscopic is meaningful to describe a model for a given system only if there is a way to look at the system in a macroscopic manner and vice versa.
- The term microscopic model is only meaningful, if there is a point of view, so that the totality of all microscopic parts of the model form at least one macroscopic object.

Usually a couple of microscopic objects are not enough to change the scale of an observer from micro to macro. This can be compared with the famous Sorites paradox (see [Kim et al., 2009]) and the question: How many rice grains does it need to form a heap? Hence, a microscopic model, whenever this term is meaningful, always consists of a high number of parts, i.e. sub-models. This finally characterises a modelling approach:

Definition 0.2: Microscopic 1

Microscopic modelling is the concept of modelling a system as summary of a high number of **sub-models**.

Applying this modelling concept, a microscopic model of the system results.

This definition is not scientifically accredited and can not be found in literature, but is the most useful for the present work.

It is vitally important to emphasise that, though consisting of sub-models, microscopic models are not developed to answer research questions that address microscopic elements: A microscopic model is a model for the macroscopic whole. Thus, the model itself might be valid

even though each microscopic element itself is not a valid image of reality. Yet, in case the microscopic elements are validly depicted compared to the real system, the summary is valid.

Therefore, the individual sub-model results can but must **not** be meaningful at all!

The dichotomy between microscopic and macroscopic is quite similar to the differences between **bottom-up** and **top-down**. A bottom-up approach starts analysing a given problem with the smallest parts (the bottom) and aggregates the findings to get information about the whole. A top-down approach starts analysing the problem as a whole (the top level) and uses decomposition to get information about the parts. Consequently, a bottom-up approach applied to model a system will lead to a microscopic model, while a top-down approach will lead to a macroscopic model. Hence

$$\text{microscopic approach} = \text{bottom-up approach},$$

$$\text{macroscopic approach} = \text{top-down approach}.$$

Yet the terms *bottom-up model* and *top-down model* are in general bad choices to describe the resulting model as top-down and bottom-up can only be used adverbial, i.e. to describe processes. On the contrary, it is worthy to discussion whether microscopic and macroscopic can be used adverbial to describe the manner a model is derived, but it seems to be the common way in literature.

The summary of all sub-models in a microscopic model is classically called **population**. Notice, that there is a subtle but important difference between the population itself, which is a set of sub-models, and the size of the population, which is a natural number. Clearly, it makes a major difference whether the population in a microscopic model changes with time or not. Addition and removal of sub-systems during run-time raises additional questions not only for the conceptional model, but also for the simulation of the model. One may think of pre-allocation of arrays and computational efforts for extension and deletion of array elements. Hence, we split the set of microscopic models into two subsets:

Definition 0.3: Population-Dynamic vs. Population-Static Microscopic Model

A microscopic model is called **population-static** if its population does not change with time. Otherwise it is called **population-dynamic**.

A population does not change if any of the sub-models changes its state, features or behaviour, as the sub-model retains its “identity”. It only changes if sub-models are removed or added.

Based on the concept of microscopic modelling many different, more distinct modelling methods have been established. Some of them will be discussed in detail in the following sections. These methods often feature a specific term that is used instead of the term *sub-model*: **Agent**, **site** or **entity** are only three of the most commonly used terms. Also **individual** is commonly used.

The concepts of microscopic modelling and simulation can (in the contrary to most macroscopic approaches) be seen to overlap heavily, as it is not easy to strictly distinguish between the formalisation of the model and its implementation. Nevertheless, we try to discriminate between a formalised model and the simulation of the model from now on.

2.1 Microscopic Modelling Methods

While differential equation models can be analysed with pen and paper, microscopic models usually require a computer to be simulated. Therefore, the history of microscopic modelling and simulation is a comparably short one. In literature two so to say founding fathers of this science could be identified: Stanislaw Marcin Ulam and John von Neumann.

According to the biography of von Neumann [Ulam, 1982] which Ulam wrote for his good friend post-hum, the two met in 1933 on a congress in Warsaw. Ulam followed von Neumann's call to Princeton in 1935 where both started a long lasting, vital collaboration. Both researchers have been invited to the, now famous but formerly secret, Manhattan-project in 1943. Undoubtedly the frenetic research for bombs and other scientifically demanding weaponry during World War II has formed the scientific basis for today's computer technology, simulation and numerics due to high availability of resources for research and the collaboration of eminent scientists of this time. Also von Neumann and Ulam benefited from the rich equipped secret research lab in Los Alamos, especially from its computer technology. Besides the progress in solving formal Physical and Mathematical problems also computer programming was stressed and improved to solve those problems, which could not be solved by hand.

2.1.1 Monte Carlo Simulation

To solve current physical problems with multi-particle systems computer technology was stressed in order to compute all particles individually while applying the average afterwards. Stanislaws Ulam established the term of **Monte Carlo Simulation** which was, to our knowledge, first published in 1951 [Ulam, 1951].

In its original concept the Monte Carlo method can be interpreted as the precursor of modern microscopic simulation. It is based on the Law of Large numbers (see Theorem 3.1), which states that the sample mean of a number of independent events converges towards the mean value. Suppose an average value of a specific random event is not known or cannot be computed, a computer can be used to sample a high number of these random events. After applying the sample mean, an approximation of the unknown value is gained.

In the 1940s Ulam and von Neumann used this technology for a couple of applications. One of them was based on ideas of Enrico Fermi in the 1930s and considers the approximation of integrals of bounded functions, in case they cannot be solved by hand.

Example 1.1: Monte Carlo Integration

Suppose a bounded, positive function f is to be integrated on the finite interval $[a, b]$. Let c denote an upper bound of the function on $[a, b]$. By definition the integral $\int_a^b f(x)dx$ is equivalent to the area between $y = f(x)$ and $y = 0$ bounded by $x = a$ and $x = b$, while $\int_a^b (c - f(x))dx$ describes the area, which is bounded by $y = f(x), y = c$ and $x = a$ and $x = b$. In sum, both areas sum up to the area of the rectangle described $a, b, c, 0$ which is $c \cdot (b - a)$.

A randomly (uniformly) generated point, sampled in the rectangle $a, b, c, 0$, clearly lies below the curve with a probability that depends on the area below the curve compared to the total area. Say a random variable $X = 1$ if the point lies below the curve and $X = 0$ otherwise, then

$$\mathbb{E}(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = P(X = 1) = \frac{\text{area below}}{\text{area total}} = \frac{\int_a^b f(x)dx}{c(b - a)}.$$

Generating N samples X_i of X, the Law of Large numbers guarantees that

$$\mathbb{E}(X) \approx \frac{1}{N} \sum_{i=1}^{N} X_i.$$

Hence

$$\int_a^b f(x)dx \approx c(b - a) \frac{1}{N} \sum_{i=1}^{N} X_i$$

poses an idea to approximate the integral.

Above version of the Monte Carlo Integration is hardly used nowadays as it can be seen to be considerably less performant than other integration algorithms (ordinary differential equation (ODE) solvers). Yet, this does not necessarily hold for high-dimensional problems. A sketch of this concept is shown in Figure 2.1 where the integral

$$\int_{0,\pi} \sin(x) \exp(-x^2)dx$$

is approximated.

Although Ulam is usually denoted as the founding father of Monte Carlo simulation on computers, the basic concepts of Monte Carlo Simulation without computers can be seen to be significantly older. Mathematician Georges-Louis Leclerc de Buffon was probably the first one who ever used a microscopic (Monte Carlo) simulation to answer a macroscopic problem: the famous Buffon's Needle Problem ([Todhunter, 1865] page 203).

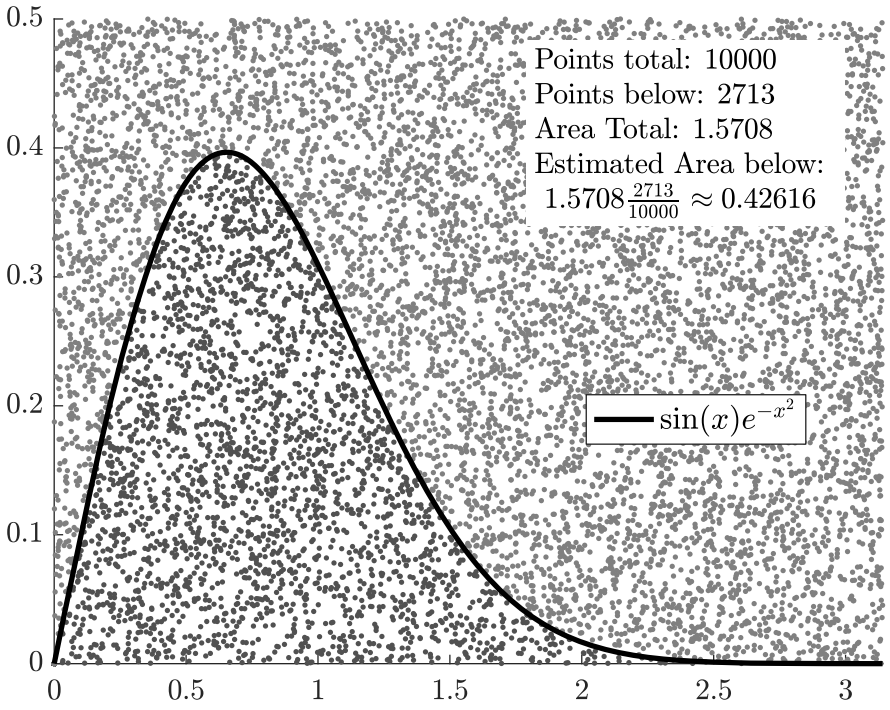


Figure 2.1: Illustration of Monte Carlo Integration on the example of $\sin(x)e^{-x^2}$. Though the function cannot be integrated analytically, the integral can be approximated by the quotient of the number of points below the curve and the number of points in total.

Example 1.2: Buffon's Needle Problem

Goal of this model is the approximation of π. Therefore, the following experiment is set up:

A sufficiently large area on a flat floor is lined. The lines are drawn vertically, parallel to each other and the gaps between the lines are a length units broad. A total of N equivalent needles with length $b \leq a$ are dropped onto the floor from some height leading to a random sample. The number of needles that intersect at least one of the lines is counted and denoted as M.

Let $X \in \{1, 0\}$ denote the random variable whether or not one of the needles intersects a line. Let $\alpha \in [-\pi, \pi)$ describe the (random) angle of the needle, then $c(\alpha) := b|\cos(\alpha)|$ describes its horizontal component. By the laws of conditional expectation (see later in Corollary 1.5)

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|\alpha)).$$

Given a specific angle α the probability that the needle intersects one of the lines is given

by the difference of c compared to a. Hence,

$$\mathbb{E}(X|\alpha) = \frac{c(\alpha)}{a} = \frac{b}{a}|\cos(\alpha)|.$$

As also the angle α is a uniformly distributed random number the expected value can be calculated by integration:

$$\begin{aligned}\mathbb{E}(X) = \frac{b}{a}\mathbb{E}(|\cos(\alpha)|) = \frac{b}{a}\frac{1}{2\pi}\int_{-\pi}^{\pi}|\cos(\alpha)|d\alpha \\ = \frac{b}{a}\frac{4}{2\pi}\int_{0}^{\pi/2}\cos(\alpha)d\alpha = \frac{b}{a}\frac{2}{\pi}\sin(\pi/2) = \frac{2b}{a\pi}\end{aligned}$$

Compare with 1.1, the sample mean M/N approximates the expected value $\mathbb{E}(X)$. Hence

$$\frac{2bN}{aM} \approx \pi.$$

A sketch of this concept is seen in Figure 2.2. A total of 1000 needles with length 0.5 were randomly dropped into a grid with line distances 1 of which $M = 312$ needles were counted to intersect. This leads to an approximation of $1000/312 \approx 3.2051$ for π. The same experiment with 100000 needles already resulted in $N/M = 100000/31827 = 3.1420$, but cannot be visualised that nicely.

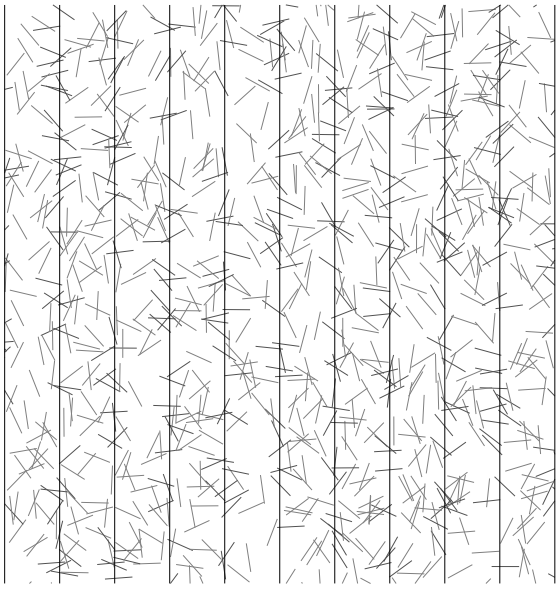


Figure 2.2: Illustration of Buffon's Needle Problem with $a = 1$ and $b = 0.5$ with $N = 1000$ needles, randomly dropped onto the grid. They are marked blue if they intersect with a lines, red otherwise.

The Monte Carlo method has been the first concept that uses bootstrapping and statistics to find answers to problems which cannot be solved directly by formal analysis. Although the single elements of the method are always stochastic (they are sampled according to distribution), the original question is usually a deterministic one. As a consequence it can be interpreted as a microscopic modelling method: the rules and complexity of the method is defined on the microscopic scale, while the results are investigated on the aggregated level.

As mentioned, the origin of Monte Carlo methods lies within (Statistical-) Physics and has been further developed to server this field of research in many ways.

- In 1953 Nikolas Metropolis, also a colleague of Ulam and von Neumann, published the famous Metropolis algorithm, which allows to generate pseudo random numbers based on Monte Carlo bootstrapping concepts [Metropolis et al., 1953]. Based on available methods for generation of uniformly distributed (pseudo) random numbers, the metropolis algorithm allows to generate streams of random numbers with arbitrary distribution. This method was refined in 1970 by W. Hastings [Hastings, 1970] and belongs to the so-called Markov-Chain Monte Carlo (MCMD) methods for efficient sampling of pseudo random numbers.

- In the 1960s, so-called Sequential Monte Carlo methods, formerly called Particle filters,were introduced. The sampled random numbers are not independent, but rely on each other sequentially. The most prominent example for this was given by Rudolph Kalman in [Kalman, 1960] by the commonly known Kalman filter.

- While the basic concept of Monte Carlo method is originally a static one, it did not take too long before the first dynamic simulation method was introduced. Kinetic Monte Carlo Method, sometimes called Dynamic Monte Carlo method, was first mentioned in 1966 by Young and Elcock [Young and Elcock, 1966]. Herein the sampled stream of random elements is defined to evolve with time according to specific transition rates.

- In 1965 McMillian laid the foundation for the development of Quantum Monte Carlo methods when he approximated the ground states of liquid helium by Monte Carlo sampling methods [McMillan, 1965]. In Quantum Monte Carlo bootstrapping and statistics are used to find solutions of the Schrödinger equation for systems of a large number of particles. In this simulation particles may evolve in time (dynamic Schrödinger equation) or interact in various ways, e.g. to match their spin or to apply forces.

- At least since 2007 the also the term Lattice Monte Carlo could be found in literature [Belova and Murch, 2007], in which particles are distributed on a lattice which allows them to make discrete jumps.

Experienced modellers might recognise parallels of mentioned methods and the “classic” microscopic modelling methods used for human-centred research problems. While physicists would denote a model as Sequential Monte Carlo simulation, a classical modeller would denote it as a Markov model. Quantum Monte Carlo simulation for multiple interacting particles would be denoted as agent-based models and Lattice Monte Carlo models could be interpreted as cellular automata. Hence, the (Statistical-) Physics community often use very similar microscopic

modelling concepts as the classical modelling community, but unites them under the term *Monte Carlo*. In contrast to standard approaches, Monte Carlo methods usually result in very specialised white-box models as the foundation for used probabilities and distributions is directly found in Physical axioms and laws (e.g. the Schrödinger equation).

2.1.2 Cellular Automata

While Stanislaws Ulam laid the cornerstone for probabilistic, continuous microscopic simulation, John von Neumann is mainly deemed to be responsible for launching the discrete and deterministic branch: The theory of (cellular) automata (see [Shannon, 1958]). Motivated by Alan Turing's Turing Machines [haz, 1988] and the idea of generating self-reproducing systems, he derived the concept of an infinite rectangular grid – each grid point has one of 29 discrete states – that, updated in equidistant steps, depicts the logic of a machine that can reproduce itself. This model, also known as the Von Neumann Cellular Automaton, can definitely be seen as one of the first **cellular automaton (CA)**. Also, the CA related technical term *Von Neumann neighbourhood* can be dedicated to this model as it was used there for the first time. Yet some earlier ideas can be found in literature for which we refer to a summarising work of Günter Schneckenreither [Schneckenreither, 2014], but their influence on the development of the modelling method was a minor one.

The concept of CAs has not become famous until the work of John Horton Conway in 1970 in which he established the famous Game of Life, which is probably still the most prominent representative of this method. First published as a "mathematical game" in [Berlekamp, 2001], the Game of Life was the first cellular automaton that became popular also among the non-scientific public.

As mentioned in [Schneckenreither, 2014] there is no unique formal definition of a CA all scientific fields agree with, yet the different definitions only differ in technical details and share the same basic components. As we are not interested in doing fundamental research on the modelling method, we only give a very shallow and rather algorithmic definition of the modelling concept. Figure 2.3 shows a standard rectangular cellular automaton before (left) and after (right) the update.

Definition 1.1: Cellular Automaton (practical oriented definition)

A cellular automaton is given by the following components.

1. **A set of cells, the cell-space.** A cell, sometimes called site, node or individual, is a passive entity that poses the foundation of the automaton. In order to establish a useful automaton a sufficiently large number of cells need to be summarised in a finite set, the cell-space. Say a number of N cells $c_i, i \in \{1, \dots, N\}$ are summarised to the cell-space
$$C := \{c_1, \dots, c_N\}.$$
2. **State and state-space.** Moreover, there a common state-space S is required. In principle there is no restriction to the state-space but sometimes a vector-space structure

is useful for analysis. Additionally, some concept that assigns each cell a specific state is required. It is commonly done via a so-called state-mapping that maps each cell onto an element of the state-space.

$$f : C \to S : \quad f(c_i) = s_i.$$

We say, sloppily, a cell c_i "has" a specific state or "is in" state s_i.

3. **Cell arrangement.** To find relationships between the cells, the cell space needs to be arranged in some way. Usually this is done by distributing all cells onto some rectangular or hexagonal lattice. This can be described formally via a bijective index mapping

$$Ix : \mathbb{Z}^K \to C : (i_1, \ldots, i_K) \mapsto c_j,$$

which assigns a tuple of whole numbers, also-called the **index**, to each cell. We say that two cells c_i, c_j are connected to each other, if $\left\| Ix^{-1}(c_i) - Ix^{-1}(c_j) \right\|_1 = 1$. The natural number K is hereby denoted as the dimension of the CA. To achieve a simply-connected arrangement, it is common to claim that for any two cells c_s, c_e a finite sequence $c_j, j = 1 \ldots M$ of cells exists so that all pairs

$$(c_s, c_1), (c_1, c_2), \ldots, (c_{M-1}, c_M), (c_K, c_e)$$

are connected.

4. **Neighbourhood and boundary conditions** In order to define some range of influence of a specific cell, a so-called neighbourhood is introduced. Thinking of a rectangular lattice, all four directly connected cells can be defined as neighbours of the cell in the middle (so-called Von Neumann neighbourhood). Moreover, we could also think of increasing the radius of influence by the four corner-cells as well (so-called Moore neighbourhood). Also, more abstract neighbour definitions could be applied. Note, that the neighbourhood must not necessarily be reflexive, transitive or symmetric. The idea of neighbourhood could be formalised via a neighbourhood mapping on the index set:

$$H : \mathbb{Z}^K \to \left(\mathbb{Z}^K\right)^J : (i_1, \ldots, i_K) \mapsto ((j_{1,1}, \ldots j_{1,K}), \ldots (j_{J,1}, \ldots, j_{J,K})).$$

Hence, any index is assigned a vector of neighboured indices. As not every cell necessarily has to have the same numbers of neighbours, J might additionally depend on the index. This is relevant especially at the borders of the cellular automaton, where the topology of the automaton is usually different than in the middle. Some extra rules need to be set up in this case.

5. **Update Rule.** The heart of any CA is the way, the states of the automaton are updated to generate a dynamic simulation method. Hereby each cell's state is updated

according to the states of the neighboured cells. Hence

$$s_i^+ = f(c_i)^+ = U(f(Ix(H(Ix^{-1}(c_i)))))$$

for some update function U. The latter is applied to all cells simultaneously. Performing the last step iteratively finally results in the dynamics of the CA.

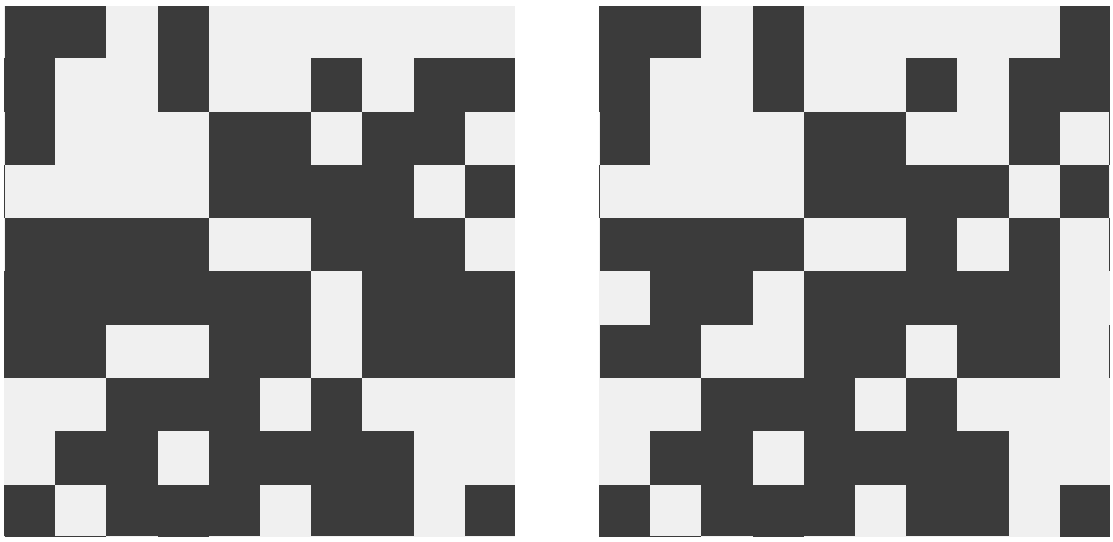


Figure 2.3: Illustration of a cellular automaton (Ising Model) before (left) and after the update (right). Cells are arranged on a rectangular grid and their states are marked with one of two possible colours. The neighbourhood used considers all 4 directly connected cells and the cell itself.

Cellular automata are known and famous for generating very complex behaviour based on simple rules. A good example for this is the so-called Ising model [Ising, 1925] shown in 2.3. This Physical model, developed from early statistical Physics, only considers two states (± 1), which describe the spin of small ferromagnetic particles. Yet it can be seen to result in very complex behaviour.

From the methodological point of view the Ising model has one feature which is not depicted in the definition of a classic CA: stochasticity. The update rule depends not only on the states of the neighbours but also on some Boltzmann-factor $e^{-\frac{J}{T}}$ dependent probability distribution.

This feature is usually excluded from the standard definition of CAs as it requires additional concepts for analysis [Schneckenreither, 2014]. Update rules of classic CAs need to be strictly deterministic. Yet many famous and interesting Cellular Automata would not be considered as such anymore due to this restriction (the Ising model would be a prominent member of this class). Consequently, it is useful to distinguish between classical CAs, wherein the same initial

distribution necessarily leads to the same dynamics, and stochastic CAs, wherein the update-rule as well as the neighbourhood might contain random elements.

2.1.3 Agent-Based Models

In 1970 economist and Nobel prize winner Thomas Crombie Schelling unwittingly laid the foundation to the development of a novel modelling approach while analysing Ghetto-ism in American cities (according to [Bianchi and Squazzoni, 2015] there have been others in the mid-1960s, however we could not find them to be sufficiently convincing). In his own words,

> *Without knowing it I was pioneering a field of study that later became known as "agent-based computational modeling",*

according to a short biography he wrote in 2005, only a short time before winning th Nobel price [Schelling, 2017] We refer to his famous *Model of Segregation* which was first mentioned in 1971 [Schelling, 1971] and analysed in more detail in 1978 in his book Micromotives and Macrobehaviour [Schelling, 1978].

Example 1.3: Schelling's Segregation Model (1)

In the 1970s Schelling investigated the behaviour of different races in the US with respect to their residence area. He observed that they form clusters of different sizes and shapes and wanted to find reasons for this behaviour. He was especially interested how the level of xenophobia impacts the size and shape of the resulting clusters. To get a picture of the underlying dynamics he decided to establish a cellular automaton for this problem. In the language of cellular automata, we could describe the model defined in [Schelling, 1971] as follows:

A set of $M \cdot N$ cells are distributed on a rectangular $N \times M$ grid. Every cell may have one of the three states: occupied by race 1, occupied by race 2 or empty. We apply the Moore neighbourhood and define a parameter $0 \leq \alpha \leq 1$. The update rule is set as follows: Say a cell is occupied with race $i \in \{1, 2\}$, then the cell is said to be *happy* if the fraction of surrounding cells with same state i is larger than α, otherwise it is said to be *unhappy*. If the cell is unhappy the cell switches state with the nearest possible cell that would make the cell happy. The rules on the boundaries of the automaton are set analogously. If the automaton remains static after one update (all cells are happy), it has terminated and the simulation is finished.

The given definition of the model is clearly that of a cellular automaton, although the switching of states with an empty cell in principle requires a more formal definition. The main reason why this model especially posed for a novel concept is not, **what** the model does, but **how** it was described by Schelling (buzzwords explicitly used by Schelling in [Schelling, 1971] marked by *italic* text):

Example 1.4: Schelling's Segregation Model (2)

Given a $M \times N$ rectangular grid. First *distribute a population* that is subdivided into two separate races. Every individual has *its neighbourhood* consisting of the 8 grid-points directly attached to it. Say that any individual that *lives in the neighbourhood* of an individual is denoted as *neighbour*. In case the ratio of neighbours that share the same race drops below α percent, the individual *moves to the nearest vacant spot that surrounds him with a neighbourhood that meets his demands.*

Although technically the same model results from both definitions, the words Schelling used to describe the model are clearly more intuitive and exceed the concepts of cellular automata. He introduced an additional layer which lies in between the cells and the state: the individual. While CAs use cells as carrier of the information of the model, Schelling's description individual that carries the state (e.g. the race) and a specific position of the grid. Hence, the cells are not the subjects of the model anymore, but become, together with the race, the objects of individuals. These individuals are henceforth denoted as **agents**, the resulting modelling method as **agent-based modelling**. This concept is depicted in Figure 2.4. The sketch shows nicely how the introduction of a new subject changes not only the inner relations of the model, but also the identity (and hereby also the number) of microscopic sub-models.

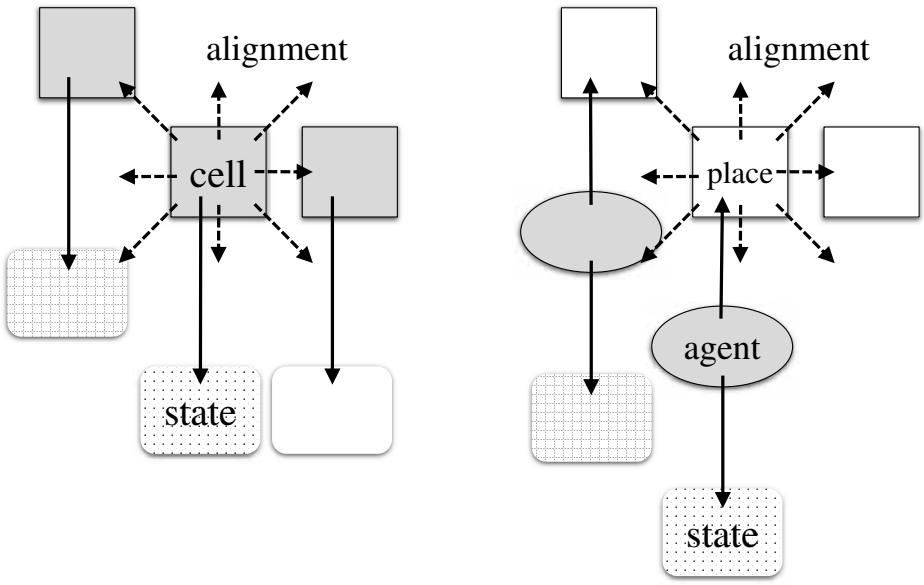


Figure 2.4: Sketch of Schelling's Segregation model depicted as cellular automaton (left) and as agent-based model. The subjects of the model are coloured with grey background.

Introducing the concept of agent to a model is not only the more intuitive and natural point of view in case of the segregation model, but it also guarantees a lot more freedom to establish

models. While cellular automata are bound to a very firm structure given by the cell space and the alignment it is very unhandy to introduce concepts like e.g.

- movement of individuals (i.e. switching of cell states as in Schelling's model),
- state-changes that depend on prior states of the cell or the automaton (e.g. a memory), or
- goal oriented behaviour.

It is moreover completely impossible to regard e.g.

- individuals that live/interact on abstract environments like networks that may change with time,
- individuals that live/interact on continuous environments,
- use different update concepts (e.g. event-driven update times), or
- uniquely track single individuals

using cellular automata concepts.

Like all microscopic modelling techniques, especially agent-based modelling requires the use of high performance computers to generate executable models. Hence, it almost took until the 1990s until agent-based modelling approaches finally became applicable for real world applications. Thomas Schelling, for example, used a chequerboard and coins to simulate his own model.

The concepts of agent-based models finally made it possible to investigate the concept of human interaction and cooperation via formal models. Probably the first one who dedicated his work to this question was political scientist Robert Axelrod. His book *Complexity of Cooperation* [Axelrod, 1997] based on his game-theoretic work *The Evolution of Cooperation* [Axelrod, 1984] has become standard literature for research on agent-based models.

In contrast to cellular automata which were in principle developed for Physical applications, the origin of so-called **agent-based models** lies within social sciences. As an inconvenient result, there is no common base that allows a formal definition of agent-based models. We already said this before about cellular automata, but the situation here is much worse. A lot of good literature is available giving ideas and tutorials to establish an agent-based model for a given question [Railsback and Grimm, 2012, Macal and North, 2006, Epstein, 2012]. Yet, it is almost impossible to state a common denominator. Therefore, agent-based modelling should **rather be seen as a general concept than a modelling method**.

We chose to stress the ideas of [Macal and North, 2006] and [Breitenecker et al., 2013] to give a broad overview about necessary and optional properties of agent-based models.

Definition 1.2: Agent-Based Model (practical oriented definition)

An agent-based model is a model that consist of (a large number of) individual sub-models henceforth denoted as agents that change with time. Every agent has a certain state, which can be split into dynamic states and static properties, and a certain behaviour that allows the agent to change its state. Hereby every agent needs to

a) be uniquely identifiable (traceable),

b) be autonomous and independent, and

c) inhabit an environment with other agents, wherein the agent needs to be able to communicate.

Every agent can additionally

d) be able to act targeted, and

e) be able to change its behaviour.

The latter two points have formerly (2006 in [Macal and North, 2006]) been stated as required properties of agents, but it turned out that it is more convenient for practical applications if these two points are relaxed.

The "definition" of agent-based modelling leaves many open questions as a lot of important points are completely left out. We give a few examples:

- The definition does not give any restrictions on **how agents should be updated**. A simultaneous update concept for all agents is common but it is also possible to update only one agent at a time while leaving the others untouched. Also, some causal update order is possible where one class of agents is updated first, while the other is updated later.
- The definition does not give any restrictions on **how time should be advanced**. We could update the model in equidistant time steps, which is the most commonly used concept, but we may use arbitrary step-sizes as well. We could also use an event-driven update.
- The definition does not give any restrictions on **how agents interact**. They may interact on the local level e.g. using some neighbourhood, some distance metric or according to a contact network. They could also interact on a global level wherein each agent uses information about the macroscopic scope of the model. Also some interaction with the environment is possible. Some definition concepts for agent-based modelling also consider the definition of "objects", i.e. static sub-models placed in the environment without their own behaviour solely generated for communication.
- The definition does not give any restrictions on **how many agents** are required to receive an agent-based model. In literature, agent-based models with one or only a few agents can be found. We do **not consider** these kind of agent-based models in this work, as we focus on microscopic models for which we required a large number of sub-models.

Consequently, the freedom for developing agent-based models is gigantic and the variety of different modelling ideas is enormous. This can be interpreted as an advantage, with respect to flexibility of the modelling method and as a disadvantage, with respect to reproducibility and documentation of a model, at the same time. We will target this problem in Section 2.3.2.

2.1.4 Microsimulation Models

In 1957 American econometrician Guy Henderson Orcutt laid the foundation to a modelling method which is commonly denoted as **microsimulation**. In [Orcutt, 1957] he stated that, although important observables for decision makers are usually macroscopic, there are a lot of mechanisms in economy that cannot be modelled on the macroscopic base. Hence, scientific decision-support in economy can only be done correctly if the microscopic level is regarded. He made this clear by a simple gedankenexperiment which is shown here with slightly modified language:

An individual decides as reaction of a given input/stimulus. Say it decides for 1 if input 1 or 2 is given and it decides for 0 if input 0 is given. Say that input is given to 100 individuals and define the sum X of all inputs and the sum Y of all outputs as the observables of the experiment. In case 50 individuals are given an input of 2 while the remaining 50 receive 0, the first 50 individuals will result in output 1 while the remaining 50 react with 0. Hence, $X = 100$ and $Y = 50$. Now consider the case that all 100 individuals are given 1 each, then each of them will respond with 1. Hence, $X = 100$ and $Y = 100$. Consequently, there cannot be a model that depicts X on Y without considering the microscopic level.

He moreover gave the ideas to a modelling method that can depict this problem:

Definition 1.3: Microsimulation (according to Orcutt [Orcutt, 1957])

The main concept originally given by Orcutt in 1957 is given as follows:

> *This new type of model consists of various sorts of interacting units which receive inputs and generate outputs. The outputs of each unit are, in part, functionally related to prior events and, in part, are the result of a series of random drawings from discrete probability distributions. These probability distributions specify the probabilities associated with the possible outputs of the unit. The appropriate probability distributions are determined by inputs into the unit and the operating characteristics of the unit. They therefore change from period to period as new inputs occur. ... Input into a unit is anything which enters into, acts upon, or is taken account of, by the unit. ... Inputs may have been produced as previous outputs of other units or they may derive from the physical environment.*

In a more formal language this idea of a model could be summarised as follows:

This model consists of a number of individual sub-models each given a set of properties (*operating characteristics*) and some initial state in a discrete state-space. The model is updated in discrete time-steps wherein each sub-model changes its state (*input*) according to given probabilities to a new state (*output*). The probabilities may depend on any kind of input from outside the system (*physical environment*), the sub-model's properties and all prior states of any number of other sub-models.

Surprisingly, Orcutt did not refer to any other work related to already developed ideas about microscopic modelling like the CA concepts of Ulam and Von Neumann. Yet the given definition can be interpreted as a very discrete version of what was described to be an agent-based model in the prior section.

Indeed, it is a very delicate problem to distinguish between agent-based models and microsimulation models. We state a few observations:

Microsimulation models, according to the given definition by Orcutt, compared with agent-based models

- cannot depict (complex) targeted and adaptive behaviour of individuals,
- are memoryless (i.e. the input only depends on the previous output – compare with Definition 4.5),
- are only updated in equidistant time-steps and hereby simultaneously for all sub-models,
- cannot take continuous states (although Orcutt mentions this as an opportunity for extensions in his work), and
- cannot use complex interaction strategies.

Contrarily, agent-based models

- are required to have at least the opportunity, to enable interaction between the microscopic sub-models.

Though the last point seems irrelevant it poses a restriction with respect to simulation of the model. A microsimulation model without interaction could be simulated in serial, i.e. one sub-model after the other, while agent-based models necessarily need to be simulated in parallel. Microsimulation models that fulfil this property are usually called Markov- or Microscopic Markov- models due to their strong relationship to Markov chains.

Definition 1.4: Markov model

A microscopic model wherein the state of each sub-model follows a Markov chain (i.e. a time-discrete memoryless stochastic process as defined later in Definition 4.5) is called (Microscopic) Markov model. The transition probabilities (transition matrix) of the Markov chain may depend on individual properties of the sub-model and external inputs.

As mentioned it is tricky to decide if a model is called agent-based or microsimulation. It surprisingly depends on the field of application up to which extent models are called the one or the other. We state some results of a small literature review on the terminology used in scientific work.

- In economy the terms agent-based (or agent-based computational economics as it is commonly called within this field or research) and microsimulation are in principle used for the same modelling concept. Papers that apply the word agent-based usually avoid the word microsimulation and vice versa. We state some examples: In [Baroni and Richiardi, 2007, Baroni and Richiardi, 2007, Basu et al., 1998, Li and O'Donoghue, 2013] the term microsimulation is used to describe the discussed model, in [Tesfatsion, 2001, Van Dinther, 2007, Axelrod, 1997, Farmer and Foley, 2009, Aoki, 2002] the term agent-based is used. Publications [Barton and Stamber, 2000, Raney et al., 2003] even defined their models as *agent-based microsimulation model*.

- Situation is very similar for models for traffic flow and pedestrian simulation. Although interaction and even goal oriented behaviour is necessary for models also the term microsimulation is used to describe related models (e.g. [Salgado et al., 2016, Wood, 2012]). Yet literature that use agent-based is the clear majority (e.g. [Ljubovic, 2009, Barthélemy and Carletti, 2017]).

- For health care applications (e.g. health technology assessment) which are closely related with population modelling mostly the term microsimulation is used [Fone et al., 2003, Spielauer, 2007, Jahn et al., 2016]. In general this is justified as most of the models addressed in the three stated literature reviews could be described by the term Markov model as defined above. Martin Spielauer [Spielauer, 2007] mentioned in 2007 that the introduction of agent-based behaviour in microsimulation models for health-care would pose for an interesting contribution to this science.

- For the modelling of diseases and disease spread, the situation is, again, different as mainly the term agent-based is used. This is obvious for infectious diseases, where human contacts are a necessity (e.g. [Lander, 2015, Miksch et al., 2011, Skvortsov et al., 2007]) but also non communicable diseases are conveniently modelled with agent-based models (e.g. [Nianogo and Arah, 2015, Li et al., 2016]).

For some concluding remarks about the terminology in microscopic models we refer to Section 2.1.7.

2.1.5 Discrete Event Models

In the early 1960s the fast developments in computer technology provoked a lot of new thinking in computer simulation. A few of the developed concepts in this decade have already been explained in the previous sections. According to [Nance, 1996] also the origin of so-called discrete event simulation can be dated to this time-period which can be seen to differ from the already introduced concepts as follows:

- Already shown concepts (CAs, microsimulation models and most of the agent-based models) are updated in discrete time steps. Hence, the dynamics of the models are defined by numbers that consider: **the amount of change per time or the probability for a change per time**.
- On the contrary, applying a discrete event paradigm a model is defined by numbers that consider: **the amount of time per change**.

Consequently, a discrete event model can be defined as follows

Definition 1.5: Discrete Event Model (analogously found in [Robinson, 2004])

A discrete event model is a model which states are defined to remain constant between any two consecutive occurrences of so-called **events**. Occurrence of such an event may lead to a state-change of the model and/or to the scheduling of new, future events.

Thus, a discrete event simulation only needs to regard the management of events, while the time in between two consecutive events is bypassed.

As a consequence, a discrete event model basically consists of two main components: A number of state variables that describe the state of the modelled system at any point in time, and a set of possible events which may lead to changes for the system variables and may lead to planning future events. Hence, simulating a discrete event model one needs to deal with two main objects: A list that stores the states of the system variables of the model at specific event-times, the so-called **state-list**, and a list that stores all planned future events, the so-called **event-list**. For simulation, one needs to run through the event list and search for the event, which is to occur in the most recent future. Furthermore, one needs to execute the event, i.e. possibly change the state variables writing into the state-list and plan new events, and finally search the event-list for the next event to occur.

The history of discrete event modelling is directly coupled to the history of discrete event simulators, i.e. simulation programmes on a computer that accept a specific description language of a model, usually a graphical one, and automatically execute the simulation. The first executable concepts have been developed in the 1960s. Among many pioneers (most can be found in [Nance, 1996]) we would like to especially mention computer scientists Keith Douglas Tocher, who was responsible for the development of the First General Simulation language (GSL) [Tocher and Owen, 2008] in 1960, i.e. the first computer simulation language that can be used to automatically simulate systems given in a specific description form applying the ideas of discrete events. A few years later Geoffrey Gordon, probably better known for his work on the General Purpose Simulation System (GPSS, an extension of the GSL) [Gordon, 1978], specifically described the use of block diagrams in order to define models for such simulation programs [Gordon, 1961].

In 1983 Lee W. Schruben introduced the **Event Graph** notation, sometimes called Simulation Graph for describing discrete event models in a graphical way [Schruben, 1983]. Later he developed the SIGMA simulator for his description language as well [Schruben, 1992].

Figure 2.5 shows a minimalist example for an Event Graph that depicts the behaviour of a bouncing ball with damping. It is a good example that the Event Graph formalism and hereby

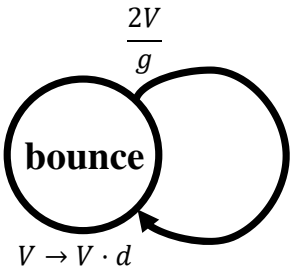


Figure 2.5: Minimalist example of a bouncing ball described by an Event Graph. The "bounce" event occurs every time, the ball hits the floor. In case of a bounce the velocity of the ball at the time instant of the bounce is reduced by a damping factor $d < 1$. The time to the next bounce can easily be calculated by the larger root of the equation $0 = \frac{-gt^2}{2} + vt$

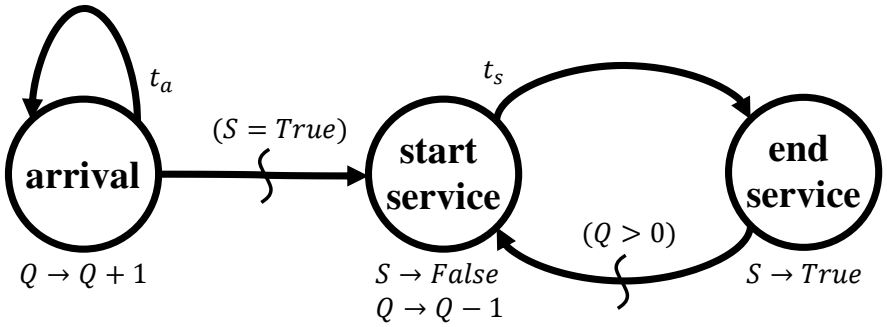


Figure 2.6: Event Graph of a single-server queue. Variable Q, initially 0, describes the length of the queue, $S \in \{True, False\}$ describes whether or not the server is available. The time between two arrivals of customers t_a and the time required to serve one of them t_s can be fixed with values or positive random numbers.

discrete event simulation can be used to depict continuous systems if one focuses on the discrete parts of the system. Yet, it turned out that this is hardly necessary for practical applications, of which most of them can be found in logistics and scheduling, summarising, so-called **process modelling**. One example is given by the Event Graph of a single server queue in 2.6 which depicts a model of a simple queuing process e.g. for a supermarket cash desk. It is clear, that depicting logistic processes like this does not require the enormous flexibility of Event Graphs. Hence, more recently developed and applied simulation tools tend to be focussed on elements that are generated (compare the arrival node in Figure 2.6) and processed through e.g. production or transportation systems (compare with the serving process on the right hand side of the Event Graph in Figure 2.6). Hereby we especially address the simulators Enterprise Dynamics [noa, c] and Anylogic [Grigoryev, 2012], but also the SimEvent library in MATLAB could be mentioned.

The resulting diagrams in these simulators are very similar to event graphs but depict the process of generated entities more directly. Figure 2.7 shows how the single server queue looks like if translated to an (AnyLogic-like) graphical description of the process.

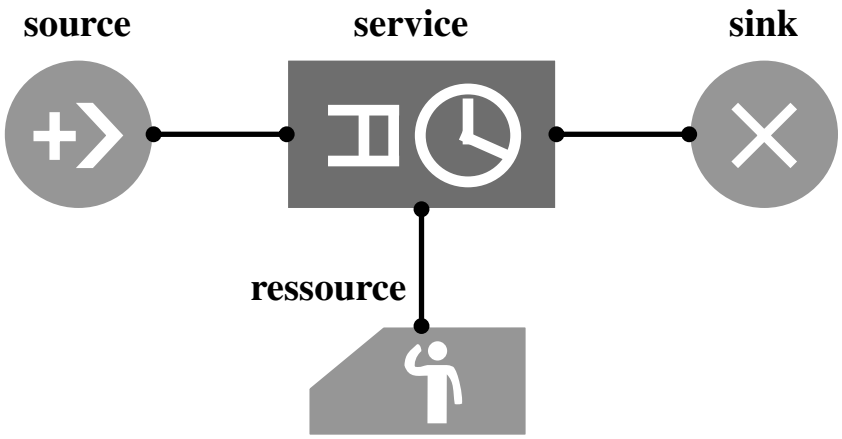


Figure 2.7: Single-server queue modelled with the Enterprise Library of AnyLogic 7. The underlying model is equivalent to the one shown in the event graph in Figure 2.6, yet the process is more intuitively depicted here.

Models depicted by such process oriented languages are always **microscopic models** although this is not generally the case for discrete event models (compare with the bouncing-ball in Figure 2.5). The dynamics of the system can easily be seen as the result of a large number of individuals, in case of discrete event simulation usually called **entities**, which are processed through the system. The position of the entity inside the process can be interpreted as its time-dependent state. Commonly used elements that are used to process the entities through the system are sources, sinks, queues, severs, path splitters and joiners, switches, and various forms of resources.

On the contrary to agent-based models, the entities are usually passive and do not "decide" about how to continue their path through the system, but are processed by their environment. This feature is probably the most obvious distinction between those two modelling approaches. Clearly, this distinction characteristic is quite vulnerable: from the technical point of view it is basically undistinguishable if the entity or its environment made the decision to proceed – both are computer programs implemented by the modeller. For this reason (discrete event) simulation expert Averill M. Law is convinced that agent-based modelling is only a special case of discrete event simulation [Law, 2007].

2.1.6 Discrete Event System Specification (DEVS)

The Discrete Event System Specification (DEVS) mechanism, developed in 1976 by Bernhard Zeigler [Zeigler, 1976], is a general mathematical concept to formally describe a model. Hence, it is in principle neither a modelling method, nor has it anything to do with microscopic or macro-

scopic models. Yet, the mechanism is very general and is, in principle, capable of depicting any kind of microscopic models. In [North, 2014] American agent-based expert Michael J. North addresses the possibility to depict agent-based models via DEVS formalism though remarking that this can only be done with restrictions:

> *Unfortunately for agent-based modeling, DEVS imposes an awkward boundary between state and behavior that makes it difficult to represent dynamically generated agent behaviors.*

He also states that there are ongoing developments to omit this problem, but in 2014 this has still been unsolved. Nevertheless, using DEVS to fully formalise a microscopic model is a great improvement compared to other usually informal description forms like decision trees, flow charts or even text or pseudo-code. Not only do we receive a reproducible model description, but it can also directly be fed to a DEVS simulator (e.g. PowerDEVS [Bergero and Kofman, 2011]).

2.1.7 Concluding Comments about Microscopic Modelling Terminology

To develop any kind of model it is essential to establish a unique and reproducible definition of the model. I.e. a description level that is sufficient for a skilled programmer to independently create an implementation of the model which produces (at least stochastically) equivalent results. In case of differential equation models this process is (said to be) comparably easy as e.g. the differential equation

$$\ddot{x} = -\omega^2 x,\ x(0) = x_0,\ \dot{x}(0) = dx_0$$

fully describes a (linearised) pendulum model. Thinking about implementation and simulation, the description is incomplete without mentioning the used numerical methods for solving the equation, but besides that, a reproducible definition is given in solely one line of text.

For microscopic simulations it is usually a lot more difficult to reach this level of reproducibility. The modeller needs to be much more careful about the definition of system variables, microscopic entities, their behaviour and states, time enhancement, initialisation and even randomness (i.e. random number generators). Hence, reproducible descriptions or documentations of microscopic models are usually long and full of technical details, which mainly disturb the reader from understanding the main mechanics of the model. There seems to be a **trade-off between reproducibility and comprehensibility** of a microscopic model description. This is a difficult problem when it comes to publishing microscopic models in journals or conferences wherein the model description should be both, reproducible and comprehensible – and, last but not least, short.

Therefore, it is vitally important to label the model or at least parts of the model with specific buzzwords that convey a certain image of the model concept. The terminus *agent-based model* will almost surely convey an image of communicating individuals with some kind of swarm behaviour. The terms *discrete-model* or *discrete event model* will clearly convey an image of a process diagram with queues and servers. Moreover, the communicated image also depends on the audience, the model is communicated to (compare with the discussion about the use of the term *Microsimulation model* in Section 2.1.4 or *Monte Carlo simulation* in section 2.1.1). Hence, it **might be formally correct** to denote a process oriented discrete event model as an

agent-based model – individuals are uniquely identifiable and (passively interact) – **but it might not be the best idea** to do so, as the term might communicate a misleading image.

From this point of view terms like *agent-based* or *microsimulation* not only label a modelling strategy which can be used to set-up a microscopic model, but they also represent a certain image that is transferred when communicating a model. Hence, we finally deduce the following three remarks.

- It is not beneficial to ask, whether a specific model **is** e.g. a microsimulation model or an agent-based model, as this question can hardly be answered uniquely. The definitions for specific microscopic model classes (even Cellular Automata) are too fuzzy and depend too much on the scientific field for making sharp distinctions.
- It is possible to ask whether a model **was modelled** using a e.g. microsimulation or an agent-based modelling approach, as there is one answer to this question only the modeller can knows. Nevertheless, the answer to this question is very likely to be useless to somebody who is interested in the model itself and not the modelling process.
- It is important to ask whether the established model **is communicated in the best way** as e.g. microsimulation or agent-based model. This question decides about the image the modeller transmits to its audience. Clearly, this has a major influence on credibility and scientific acceptance of the model and it needs to be done with care.

Finally, the discussion about problems with communication of microscopic models in terms of the lack of unique definitions is another motivation for finding a formal classification that **can** be used to uniquely label a microscopic model.

2.2 Simulation Methods for Microscopic Models

The lack of a firm modelling method definition is both an advantage (mentioned in 2.3.2) and a weakness of microscopic models. The huge level of flexibility basically all microscopic modelling methods provide makes it very hard to develop simulation environments that are capable to guarantee this flexibility to its full extent. Hence, most simulators for microscopic models are quite restrictive. As a result, mainly two simulation strategies can be found:

- The model is implemented from scratch using a fundamental programming language e.g. by using loops and if clauses.
- The model definition is (re)formulated so that it can be used in a microscopic simulator with large restrictions.

On the contrary to partial- or ordinary- differential equation models, the first strategy based on "hard-coding" is comparably easy to accomplish, as it is not necessary to make use of difficult numerical algorithms (e.g. like ODE solvers, finite element method or advanced zero crossing methods). Microscopic models can usually be implemented quite straight forwardly and only require the basic elements of almost any programming language. Yet other challenges occur which will be stressed in Section 2.2.1.

In case one aims to use one of the predefined simulators for microscopic models (a few examples will be given in 2.2.2), it might be necessary to adapt certain parts of the model to the restrictions of the simulator which might be quite demanding. Hence, it quite often happens that implementation in the simulation seems to work quite well and fast at the beginning, but might still be destined to fail in the end, as specific, delicate parts of the model are surprisingly difficult to implement due to restrictions.

In recent years alternative concepts apart from the two mentioned strategies have been established. We already mentioned the idea of modularity of microscopic models, wherein it is intended that specific parts (modules) of the simulation or at least a validated base simulation can be reused and easily coupled to establish new executable models. This concept is one of the main objectives of research project DEXHELPP. This FFG-financed project aims to improve the Austrian health-care system by data- and model-based decision-support. Hereby a validated model for Austria's population poses the base-level, while models for morbidity, health-care demand and supply, intervention scenarios or payment system are conceptualised as modules on top [Bicher et al., 2015]. (More about this specific model is given in Section 5.6.1). The first publications that prove that this is more than only a theoretical concept, are going to appear in near future as the mentioned population model has already been enhanced by three different modules for health-care research purposes [Zauner et al., 2017]. Anyway, modularity of models is very restrictive with respect to applicability, but it is very efficient in case a greater number of models for similar purposes are required.

Additionally, we would like to mention the concept of multi-level modelling languages for microscopic models [Warnke et al., 2015]. This idea is also developed for efficient development of a large number of models for similar purposes. Yet, instead of complete simulation parts, small sub-processes that have a prominent role in the hard-coded simulation model are combined to much semantically shorter commands. Hereby some kind of meta-language is generated that allows to quickly establish new models for a specific purpose without really taking care about technical details. For an agent-based population model, implementation of the hard-coded operations: *Create a new agent-instance, extend the agent pool, assign sex and a unique ID, assign age* 0*, update the family record,* ... could be summarised to a much shorter and more representative *give-birth* command.

2.2.1 Microscopic Simulation - Hard Coded

Whenever a modeller/programmer wants to exploit the full flexibility of microscopic modelling it is unavoidable to use a fundamental programming language. Hereby the modeller/programmer accepts the problem, that probably more time is spent on the implementation of the code, than using a predefined simulator. We distinguish between procedural and object-oriented programming languages. The latter are clearly the more advanced and much younger. As for several other applications (like web programming etc.) they are commonly said [Axelrod and Tesfatsion, 2006] to be advantageous for microscopic model implementation.

This trend is a consequence of the smaller difference between the way microscopic models are defined (i.e. as a conceptional model) and the way they can be implemented. Object oriented programming languages make it possible to use the microscopic sub-model (entity, agent, cell, ...) to be the centre of the simulation as so-called objects (to be precise instances of objects).

Hereby a clear distinction between the implementation of the microscopic parts of the model and the, superior, "macroscopic" system of the model, which takes care about how the microscopic elements work together, is given. This is not possible for procedural languages wherein the sub-model implementation needs to be embedded in the implementation of the macroscopic system. Mentioned distinction between the implementation of the microscopic and the macroscopic behaviour of the model is not only more intuitive when translating the conceptual model to hard code, it is also more flexible with respect to model changes, improves readability of the code and, therefore, positively influences the verification and validation process. Figure 2.8 shows the difference between a microscopic model implemented using a procedural and an object oriented programming language on the example of a population model.

Of all object-oriented programming languages Java [noa, d] and Python [noa, e] are clearly the most prominent with respect to microscopic modelling (and probably in general). They are not only extremely flexible when implementing microscopic models, they are also unproblematic regarding software licenses – they are essentially free-ware. The latter is another bonus of accepting the additional workload when hard-coding a microscopic model in a rather fundamental programming language in comparison to using pre-developed simulators.

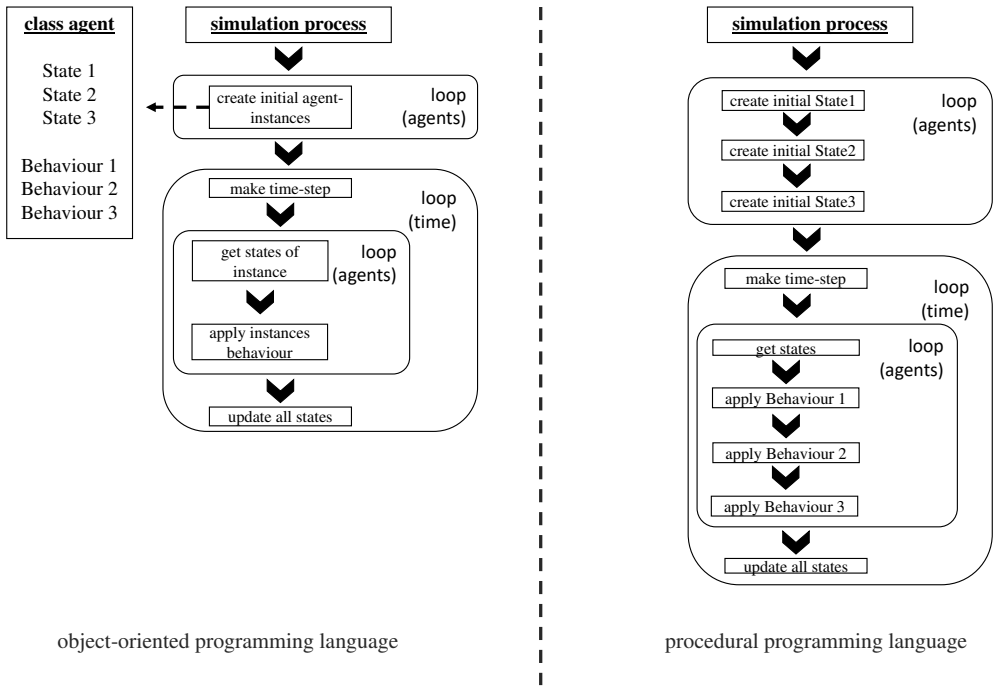


Figure 2.8: Microscopic model (agent-based model) implemented in an object-oriented programming language (left) compared to the same implementation in a procedural language (right)

2.2.2 Microscopic Simulators

Existing microscopic simulators or simulation environments have a couple of properties in common. First, as mentioned before, they are quite restrictive regarding how the conceptional model is defined and which modelling concept was used to develop the model. Second, they are usually limited with respect to the number of individuals, due to large computational overhead (e.g. because of a graphical user interface). For this reason, most microscopic simulators are mainly used for prototyping, while the final applied model is developed in a more fundamental language.

In this section we want to summarise a couple of prominent microscopic simulators and state a few of their features.

- **AnyLogic [Grigoryev, 2012, noa, a]:**
  The Java-based simulator AnyLogic is probably the most flexible microscopic simulator. Besides the macroscopic modelling technique System Dynamics, AnyLogic provides graphical environments for agent-based modelling and discrete event modelling in the process oriented sense (Enterprise library). The latter is easily applicable and offers great flexibility with respect to different interarrival distributions of entities and entity attributes as well. The agent-based features of AnyLogic are a little bit more difficult to apply as they require Java programming skills. Yet, the most important feature of AnyLogic is to allow the combination of all three modelling methods to establish hybrid simulations. Unfortunately, large overheads due to the rich graphical user interface prevent simulation of models with a really large number of sub-models. Based on experience the limit is reached with about 10^5 individuals for agent-based models with mediocre demanding behaviour and features.

- **Netlogo [Tisue and Wilensky, 2004, Wilensky,]:**
  Based on personal experience the simulation language NetLogo is currently the best simulation tool for agent-based models and cellular automata. It is a development of the Logo language, which became popular in the 1980s for teaching purposes and combines a specifically developed programming language (comparable with the already mentioned concepts of multi-level modelling languages [Warnke et al., 2015]) and a simulation tool that automatically simulates and visualises the simulation results. As it is a free-ware product and the developers are open to cooperate with scientific partners, a lot of research is going on to improve the quality and performance of the tool. Currently the tool is mainly used for educational purposes, but the software and the programming language will probably gain in impact during the next years due to these developments.

- **Arena [noa, b] and Enterprise Dynamics [noa, c]:**
  Both, Arena and Enterprise Dynamics, are classical (microscopic) discrete event simulators which can be used for the modelling of processes. They are, both, highly advanced and are equipped with a great number of toolboxes for specific fields of application. They additionally offer a user-friendly environment for 2-d and 3-d visualisation of the simulation process.

- **TreeAge [noa, f]:**
  TreeAge is the most frequently used tool for simulation of microsimulation models in

health-care, which, as mentioned in Section 2.1.4, are essentially microscopic Markov models. It has become the “holy grail” for modelling health-technology-assessment during the last decades. Moreover, it is known for its ability to intuitively depict state-discrete Markov processes, not by their transition matrix, but by a transition graph (the ”tree” according to the User’s Manual available from the web-page cited above). Herein every node corresponds to one of the states while a directed weighted edge between them indicates that a transition between two states is possible and how high the probability for it is. For its simplicity and the intuitive graph-based structure TreeAge became one of the most popular tools for researchers without or with hardly any mathematical background.

Clearly, there are hundreds of other microscopic simulation environments, but the selected ones were detected to be the most general ones with respect to the field of applications. Most other simulators are specialised in implementing models for specific research areas (e.g. traffic modelling, fluid flow modelling, population modelling,...).

2.3 General Properties of Microscopic Models

French Physicist Eric Bonabeau is one of the leading researchers in complexity theory and swarm intelligence. In one of his articles about agent-based modelling he stated three general properties of microscopic models that distinguish the modelling approach from most other modelling techniques (see [Bonabeau, 2002]). As agent-based modelling is clearly the most general microscopic modelling method these three statements hold for microscopic models in general, if interpreted correctly.

- Microscopic models can depict *emergent behaviour*.
- Microscopic models are flexible with respect to model extensions and modifications.
- Microscopic models pose for a natural description of the system.

In the following we will critically discuss each of them in more detail.

2.3.1 Emergent Behaviour

In general a system is said to generate so-called **emergent behaviour** if the observed behaviour is in some way surprising or unexpected. Consequently, it is a very philosophical question, which behaviour is really “emergent” as it might surprise the one, while being a logic consequence to an other. To discuss this topic more formally, we cite American biologist and complex system scientist Peter Corning who stated (in [Corning, 2002]) the following characteristics of emergent behaviour, based on the work of Jeffrey Goldstein [Goldstein, 1999] and more extensive literature research:

> *The common characteristics (of emergent behaviour) are:*
>
> *1) radical novelty (features not previously observed in systems);*
>
> *2) coherence or correlation (meaning integrated wholes that maintain themselves over some period of time);*

3) a global or macro level (i.e. there is some property of wholeness)

4) it is the product of a dynamical process (it evolves); and

5) it is ostensive (it can be perceived).

In 1) Corning addresses the idea, that behaviour which is generally labelled as *emergent* is somehow new to the observer and has not yet been observed in this system. Otherwise it would not be unexpected anymore. This also matches 5) as emergent behaviour is nothing you specifically need to look for, but, also using the words of Corning ([Corning, 2002] page 22),

you will know it when you see it.

Parts 2), 3) and 4) are dedicated to the causes of emergent behaviour and are probably the most interesting for a modeller. They address that emergent behaviour occurs if there is an observable *whole* that changes dynamically with time. As this *whole* can be interpreted as a macro-level, also a micro-level needs to be included. Consequently, emergent behaviour occurs for the macro-level of system consisting of microscopic elements.

Example 3.1: Bird Flocks

The most obvious example for emergent behaviour in systems can be found directly in nature: bird flocks. A flock of starlings is known to consist of up to tens of thousands of individual birds, yet seen from afar, the complete swarm poses as a gigantic tube which murmurates comparable with the bubbles of a lava-lamp. This object and its unexpected sudden movements are probably the most beautiful example for emerging behaviour: it is obvious (ostensive) and unpredictable. The behaviour of the swarm is (solely) determined by the behaviour of the microscopic elements of the system, the birds. Their behaviour on the individual level, as simple as it is, results in highly complex behaviour for the total swarm due to interactions of tens of thousands of individuals.

One of the key features of microscopic modelling is the possibility to easily **depict** such emergent behaviour. Hereby the key term in the sentence is “depict” as this feature is quite often misunderstood:

It is a general property of microscopic models, that emergent behaviour can be observed in model output. Yet, this alone is **not** the advantage of microscopic modelling: It is in general unwanted if one is surprised by emergent model results as it makes the validation process more complicated. It **is** a feature of microscopic modelling that a microscopic model it is easily able to depict emergent behaviour if it is observed in the real system.

Example 3.2: Bird Flocks - Boids Model

According to the given bird flock system (Example 3.1) artificial-life expert Craig Reynolds published his famous Boids-model in 1987 [Reynolds, 1987]. As implied, the motivation for this microscopic (in this case, agent-based) model was originated in the observation of a bird or fish swarm. The individual birds are modelled with three very simple rules, namely, get attracted from the centre of the swarm, adjust your direction miming your neighbours, keep a safety distance to your neighbours. Yet the swarm, seen as a whole, behaves, similar to the real system, almost arbitrarily complex. Hence, the emergent behaviour of the real system is depicted in the model. A sample screenshot of a 3-d Boids simulation can be seen in Figure 2.9.

It is worth to discuss whether the emergent behaviour of a real bird-swarm is depicted validly or not. Nevertheless, the model was not developed to explain **how** a bird swarm moves in the sense of a prognostic model, but **why** it moves in that unusual and unexpected way, and **that** these simple rules on the microscopic level may lead to almost arbitrarily complex behaviour of the swarm. From this point of view, it is clearly a valid model.

Although commonly believed, the possibility to depict emergent behaviour, is (in our opinion) not only a property of microscopic models alone. One famous example for a macroscopic model that shows emergence was developed by American mathematician Edward Lorenz (who is commonly believed to have introduced the term *butterfly effect* in chaos theory). His nonlinear ordinary differential equation system

$$\frac{dx}{dt} = \sigma(y - x), \tag{2.1}$$

$$\frac{dy}{dt} = x(\rho - z) - y, \tag{2.2}$$

$$\frac{dz}{dt} = xy - \beta z \tag{2.3}$$

published [Lorenz, 1963], also-called Lorenz attractor, is known to result in arbitrarily complex, almost chaotic solutions for specific parameters and initial conditions, which clearly seem "emergent" on the first sight. A sample path is depicted in Figure 2.10. Since 1963 the fundamental processes behind the attractor have been rigorously studied e.g. using bifurcation analysis and the main why- and how- questions behind the strange paths of the solution curves have been answered satisfactory. Hence, many mathematicians would not consider the behaviour of the solution curves as emergent nowadays – the magic has disappeared. Unanswered we would like to raise the question: Would we still consider the behaviour of the Boids model as emergent, if we analysed it as rigorously as the Lorentz attractor?

Besides from the Lorentz attractor, which was mainly derived for academic purposes to show the concept of chaos theory, also the Navier-Stokes equations can be stated as example for emergent behaviour in macroscopic models. The occurrence of turbulences in their solutions is mostly emergent (e.g. [Li, 2007]).

Consequently, it is, in our eyes, not necessary to model a system in a microscopic way in order to depict emergent behaviour, yet it is surely the **easiest way to do so**, as microscopic mod-

els are defined on the microscopic level while the results are observed on the macroscopic level. This can be interpreted as a natural way of disguising the actual mechanics that are responsible for the output. We are easily able to generate models that show some version of emergence, but we have no guarantee that the specific behaviour the model shows can be found in reality as well, or if it is just a by-product of modelling errors.

This has not been a problem for the Boids model as the research problem for this model was very shallow, but it might be a problem when it comes to applications where we not only aim to find more meaningful qualitative but also quantitative model results to perform prognoses or scenario tests for real-world problems.

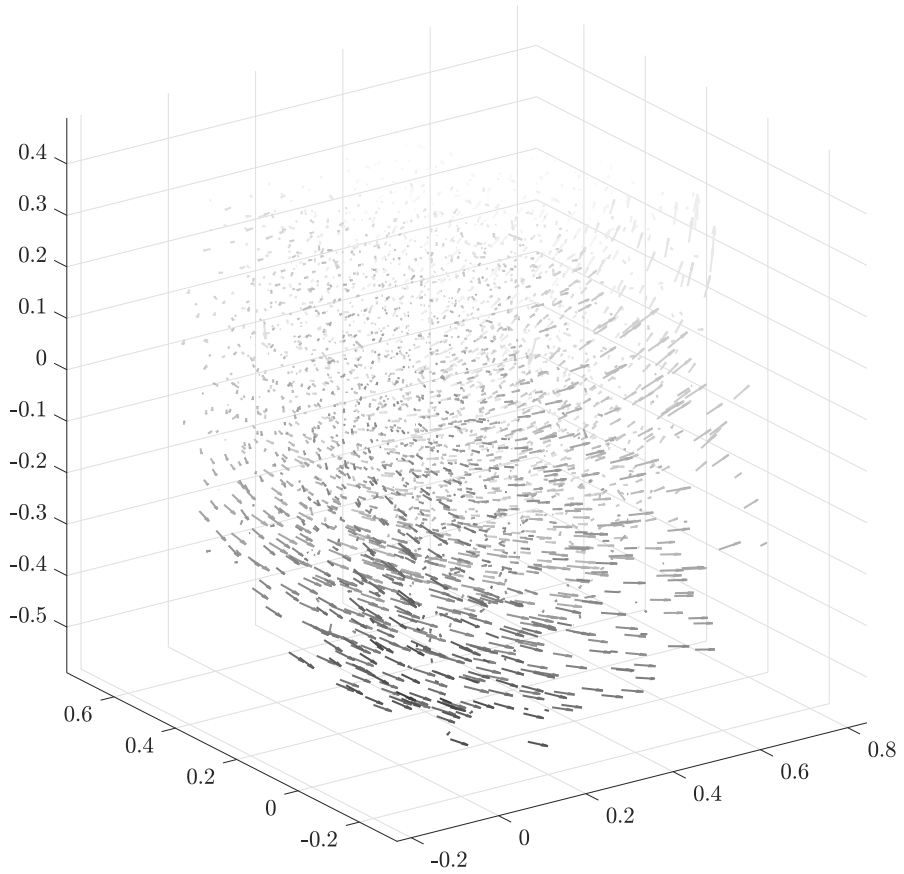


Figure 2.9: Screenshots of a 3-d Boids simulation run with 4000 individual swarm members. While the single sub-models only follow three simple rules, the swarm, as a whole, moves almost unpredictable.

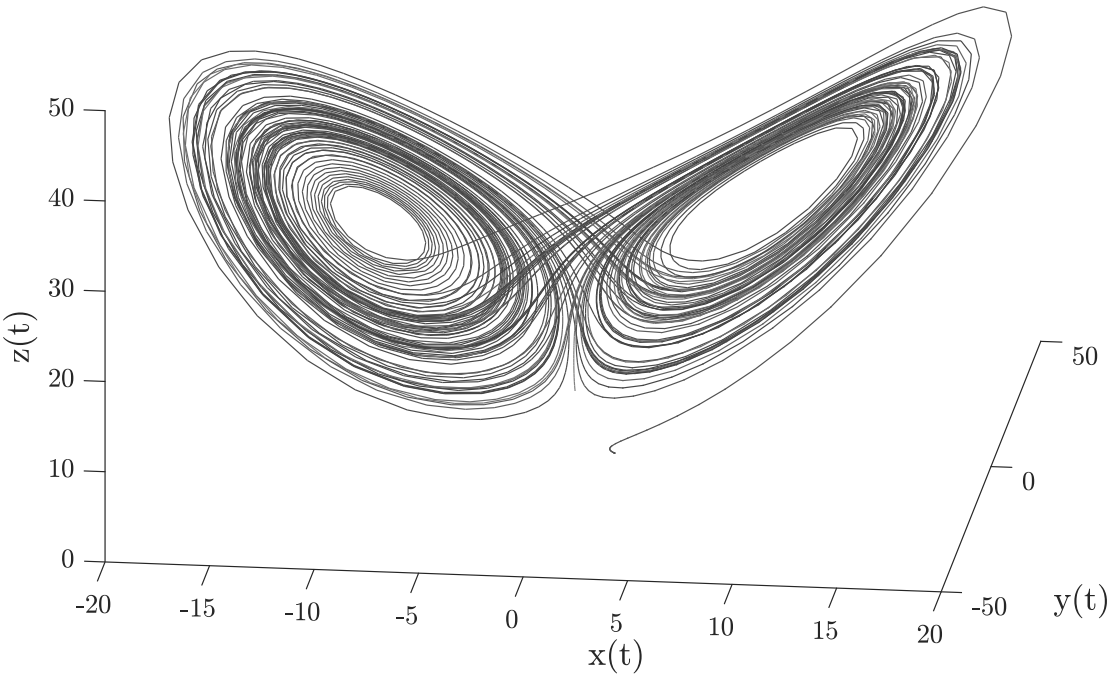


Figure 2.10: Sample solution of the Lorenz attractor for $\sigma = 10, \beta = 3, \rho = 28$ in three dimensions x, y, z. The curve runs from blue $t = 0$ to red $t = 100$.

2.3.2 Flexibility

It is a matter of our ephemeral times that the amount resources and time modellers are given to develop new models to given research problems is steadily decreasing.

On the one hand, this is justified as new technologies definitely make model development easier. It is the concept of so-called **simulators**, that models do not have to be implemented totally from scratch with a fundamental programming language like C or Fortran as specific basic modelling parts are already pre-implemented. Moreover, also much more user-friendly object-oriented programming languages like mentioned Java [noa, d] or Python [noa, e] have become almost equivalently powerful and are continuously extended by modules that make the life of programmers considerably easier. Additionally, it has become a lot easier to get access to necessary validation data (as discussed in Chapter 1).

On the other hand, with the increasing availability of data as well as more and more complex research questions also the requirements on the quality of models increased. Summarising, more

and more complex models need to be developed validly in shorter time.

As a result **reusability** of valid models or at least of valid model-parts for different research questions has become vitally important. This idea can be enhanced even further by developing single so-called **modules** which are investigated stand-alone small valid models themselves and may be linked together to form larger combined models. Consequently, the question arises:

Given a research problem where this question makes sense: is a macroscopic or a microscopic model more capable for developing reusable modules?

In almost all circumstances the answer to this question will be the microscopic approach. Reason for that is the enormous level of **flexibility** the microscopic model offers compared to macroscopic models for the same problem. Clearly, flexibility is the key to modularity as it is necessary to define interfaces to couple and to extend the model. We give a few ideas why microscopic approaches, probably with exclusion of classic cellular automata, are usually more flexible than analogous macroscopic models:

- First, very generally, the fact that microscopic models are defined on the microscopic level makes it possible to extend models on the microscopic scope, but also on the macroscopic scope. Note, that although the dynamics of a microscopic model are usually given by the microscopic elements, it does not exclude that additional dynamics are defined on the macroscopic, i.e. the aggregated level. Clearly, macroscopic models can only be extended on the macroscopic level – **aggregation** of microscopic elements **is always easier than decomposition** of macroscopic ones.

- Secondly, the definition of microscopic models is a lot less formal than that of macroscopic ones. Hence, it is a lot easier to add and to change specific components. Usually some state- or flow-chart-like notation is used to specify the behaviour of the sub-models which is easily adapted and modified. Macroscopic modelling methods, on the other hand, usually require (differential-) equations and, hence, a more **mathematical** and less **algorithmic notation**. Also, boundary-, initial- conditions, state-space, and state-variables always need to be defined formally.

- Thirdly, a lot of model extensions for systems with a micro-macro-scope regard the introduction of (additional) heterogeneity on the individual level. One may think of introducing a disease in a population model (persons may be diseased or not), additional species in a cohabitation model, different individual biological fitness levels in an escape routing model, or different types of cars in a traffic model. On the microscopic level, these ideas pose for a very easy model modification as only one additional variable needs to be introduced. On the macroscopic level a so-called **state-space blow up** effect can be observed: Each additional characteristic introduced to the model, say an individual may have this characteristic or not, the total number of (differential-) equations to describe the system on the macroscopic level is multiplied by two.

We give a small example which specifically emphasises the last point.

Example 3.3: Case Study: Agent-Based Versus System-Dynamics

As the main result of project GEneric POpulation Concept (GEPOC) in the course of mentioned research project DEXHELPP we developed two different versions of a population model for Austria, a microscopic and a macroscopic one [Bicher et al., 2015]. On the one hand, this idea made it possible to use either of the models to verify the functionality of the other. Also, some cross-validation attempts were done. On the other hand, it is possible to decide whichever of the two models is better suited for a model-extension to a specific application. The first model is a classic agent-based model with birth, death and migration processes. Age and sex are were defined as states of the agent. It is, in more detail, explained in Section 5.6.1.

The macroscopic model was developed using system dynamics. The concept of system-dynamics, introduced by J.W. Forrester and first published in 1961 [Forrester, 1961], can be stated as the most flexible and modular concept to macroscopically model and simulate black-box systems that consist of microscopic elements. It is based on ordinary differential equations which are well hidden behind graphical stock and flow diagrams comparable with blocks and connectors in standard physical simulators. A stock represents a total number of a certain type of microscopic elements, while the flows depict their change in time. The formal equation basis is automatically generated and simulated using supporting simulators. System dynamics models can be enhanced easily by introducing new flows and stocks and can, therefore, be denoted as the only serious alternative to microscopic simulation models with respect to flexibility.

In the GEPOC system dynamics model the population was split into a high number of stocks. As a fine age-resolution was necessary, 99 stocks had to be used to simulate the change of population in each age-class. As sex was an additional layer of heterogeneity, a total of $99 \cdot 2 = 198$ stocks define the model. It was simulated using the simulator Anylogic [Grigoryev, 2012, noa, a]. As seen in [Bicher et al., 2015] the system dynamics model worked reliably and even turned out to be a lot faster than the agent-based approach.

To compare the level of flexibility, we introduce a hypothetical research question: *We aim to compare the total number of patients suffering from diabetes in specific NUTS3 regions in Austria.*

Clearly, each agent in the agent-based approach needs to be extended by two additional states. First the agent needs to be assigned a specific residence (district) e.g. via a distribution, second a diabetes-state needs to be introduced with specific incidence rates.

In the system dynamics model each population stock needs to be subdivided into 35 (NUTS3 regions in Austria) stocks of which each of them, again, needs to be subdivided into 2 (diabetes: yes or no) stocks. Consequently, a system-dynamics model with $2 \cdot 99 \cdot 35 \cdot 2 = 6930$ stocks results, which is neither executable nor verifiable anymore.

We could blame the modelling method for not being suited for the given research question, but this would not be totally fair: A system dynamics model that only considers diabetes cases and districts (as specified in the research question) with $79 \cdot 2 = 158$ would be possible. The main problem lies within the ineligibility of the underlying population model as a modular and flexible basis model. Consequently, only the agent-based model has been used for applications in health services research so far.

Summarising, microscopic models usually turn out to be very flexible with respect to model extensions and coupling. Herein they usually outperform their macroscopic alternatives for the given research question.

2.3.3 Natural Description of the System

It is commonly known that there is always some gap between researcher and decision maker which makes communication difficult. This is a tricky problem that needs to be overcome to perform scientific decision-support. This can be seen by investigating the vast number of scientific publications solely dealing with this topic (some examples [Verdon-Kidd et al., 2014, Uzochukwu et al., 2016, von Winterfeldt, 2013]). They basically denote the question *how to communicate with decision makers* to a science of its own. This gap is especially large in human-centred sciences like epidemiology, sociology, biology or economics wherein most decision makes are either politicians or experts in the specific field with hardly any technical background, especially regarding mathematics or informatics.

Most of applications for microscopic modelling can be found in mentioned sciences. Hence, this problem is very present. Independent of the used modelling approach and validity of a developed model, decision makers will not deem the model as **credible** if

- the model is not internationally accredited as a valid standard-method (e.g. by a high number of ranked publications) or
- they do not at least get the impression of understanding the concepts of the model.

Regarding the first point, we already discussed in Section 2.1.7 that there are difficulties to publish microscopic approaches as reproducible and comprehensible definitions of models are very hard to establish, especially when publication-volume is restricted.

When required to develop new methods for a given research question, the first bullet point is no option anyway. To satisfy the second discrepancy, it is the modellers responsibility to give the involved decision maker a basic image of the model functionality to arouse credibility. Hence, if comparing a microscopic with a macroscopic approach for the same research problem, the question arises, which of them is easier explained to persons without modelling background. We will henceforth denote these persons short as non-experts, indicating that they are not experts for modelling or any other science which is based on mathematics or informatics. This should not be confused with the idea that they are not experts for their original research field.

As mentioned in the previous section microscopic models are usually defined in an algorithmic way using state-charts or other causal diagrams that describe the behaviour of the sub-models. Compared to macroscopic models which are usually defined via (differential-) equa-

tions, these processes are, in principle, a lot easier to explain to non-experts. Though this is obvious, it cannot be used as an argument for prioritising the use of microscopic models to macroscopic ones for two reasons: First, there are macroscopic approaches, especially system dynamics, that are specifically developed for simple communication of macroscopic models. Secondly, it is not necessary and sometimes not even beneficial to communicate the whole model mechanics to convince a decision maker of its credibility.

Modellers [Waldherr and Wijermans, 2013, Jahn et al., 2014] as well as visualisation experts [Parry, 2011] agree that the key to communicate simulation models to non-experts lies within graphical representation. Herein microscopic models are clearly more suited than a macroscopic analogue:

- First, microscopic models can be visualised in a lot of different representation forms: They can be visualised on the microscopic or on the macroscopic level and any level in between. Moreover, visualisation is not bound to representation of the simulation results but can be used to show the dynamics of the simulation during runtime on any layer as animations. Macroscopic models can only be shown on the macroscopic level and animation does not really improve understanding of the results.

- Secondly, visualisation of the model on the microscopic layer directly shows the viewer what are the sources for the aggregated behaviour of the model, while these processes are hidden behind formulas in macroscopic models. Humans are used to observe their environment analytically and, therefore, on the microscopic level. Hence, they are more qualified to decide about the validity of a presented process if they can see their components and how they work together.

The concept of *causality* additionally supports the last point. Clearly, the causal reason for the behaviour of the aggregate follows is given by the individual parts. If we do not directly regard the individual parts by modelling the system macroscopically we cannot see this causal relationship. Hence, the microscopic approach is referred to as the **natural description of a system** by Bonabeau [Bonabeau, 1997]. All arguments that support the use of microscopic models in this section can, in principle, be attributed to this fact.

CHAPTER 3

Stochastic Methods

As mentioned in the introduction, microscopic models are usually applied for highly complex systems wherein even the microscopic elements show a tricky behaviour. Consequently, models for individual sub-processes of microscopic models are usually black-box models and, as it is usually the only feasible approach to model them, contain random elements. Hereby it is irrelevant whether the real system components behave randomly or are simply too complex to describe.

Simulation of human behaviour is the most obvious example for this: We want to model the behaviour of a person that needs to pick one of two possible options. In reality the human decides based on personal memory, observations, logic, feelings and a variety of other factors, but very rarely purely on randomness (e.g. by a coin-flip). Thus, the process would be considered as deterministic in reality. Nevertheless, it is basically impossible to model the process deterministically in a simulation model for one of these reasons:

- The real process of making a decision requires components which are not included in the model.
- The real process is not fully known.
- The real process is too individual.
- To include the complete complexity of the real decision process would make the model unusable as it cannot be parametrised

As a result microscopic models are usually stochastic models and require different treatment than deterministic models. Hereby we especially address parametrisation, calibration (identification of parameter values), sensitivity analysis, simulation output interpretation, result post-processing and, in general, any kind of formal model analysis.

As a lot of the mentioned features are focus of the current work this chapter to gives quick summary of basic formal stochastic methods which will be needed to analyse stochastic microscopic models. As stochastics can be split into probability-theory and statistics we first discuss

a couple of basic theorems and definitions for these two research areas. Then we discuss some applications of probability theory for sums or random numbers via limit theorems, which are very interesting for aggregated analysis and result post-processing of microscopic models. Later we introduce the concept of stochastic processes.

All theorems and definitions presented in this section can be found in standard literature for statistics, measure and probability theory. Thus, we do not state any proofs. For more information the reader is referred to literature, e.g. [Gardiner, 2009].

3.1 Probabilistic Basics

3.1.1 Probability and General Concepts

The letter P indicates a so-called **probability measure**, i.e. a positive measure normed to 1. We write

$$P(A) := P(\text{an event in } A \text{ occurs}) := \text{probability that any event in } A \text{ occurs.}$$

Interpreted as a generalised function that maps events to correspondent probabilities we also denote P as a (probability) **distribution**. Moreover, a **probability space** is given by a triple $(\Omega, \mathcal{A}, P)$ wherein $\mathcal{A}$ denotes a σ**-algebra** on Ω, the so-called **sample-space**, and $P : \mathcal{A} \to [0, 1]$ is a **probability measure**. We conventionally write

$$A_\sigma(\Omega) = \mathcal{A}$$

whenever a sigma algebra is defined on a space Ω in order to save precious variable names and to instantly show the direct relation between σ-algebra and underlying space. It is only used when the choice of the σ-algebra does not matter or is clear, given the context. It should not be confused with the sigma operator $\sigma(\mathcal{M})$ that defines the smallest possible sigma algebra containing all sets in $\mathcal{M}$.

When considering integrals of measurable functions $f : \Omega \to \Gamma$ we interpret

$$\int_\Omega f(x) d\mu(x)$$

in the Lebuesgue sense for $f : \Omega \to \mathbb{R}$.

The calculus of the Lebuesgue integral is highly influenced by the properties of the underlying space and especially if it is **continuous** or **discrete**. These two are commonly used terms to describe two contradicting fundamental properties of a set or space. Surprisingly they are very rarely formally defined in literature, as they are rather self explanatory. Nevertheless we deem it important to give a definition here, as the border between continuity to discreteness is for one of the central topics of this work.

Definition 1.1: Discrete Set

A totally ordered set A is called discrete if for any element x both sets $\{y \in A : y < x\}$ and $\{y \in A : y > x\}$ are either empty or have a maximum and a minimum.

Above definition surely leads to a countable set – i.e. the elements can be enumerated and there is a bijective mapping from A to some subset of $\mathbb{N}$. In most applications for modelling and simulation discrete is usually equivalent with **finite** as a computer, responsible for the simulation, cannot deal with infinite sets anyway. In this case every function defined on a discrete set A can be represented via a finite lookup table which can be written as a $|A|$-dimensional vector. Consequently, functions on $A \times A$ can be written as $|A| \times |A|$ matrices.

On the contrary, functions on continuous sets need to be described with analytical means.

Definition 1.2: Continuous Set

A totally ordered set A is called continuous if for any element $x \in A$ the set $\{y \in A : y < x\}$ has no maximum and set $\{y \in A : y > x\}$ has no minimum.

The definition above implies that in between any two elements $x < y \in A$ we can always find a third element $z \in A$ with $x < z < y$. Consequently, the number of elements $\{z \in A : x < z < y\}$ can never be finite and any continuous set always has an infinite number of elements (note that a set with only one element cannot be continuous). Thus, it is (in general) impossible to describe functions on continuous sets via tables, vectors or matrices, but we require algebraic expressions.

Remark 1.1:

Note, that we do not distinguish between $\mathbb{R}$ and $\mathbb{Q}$ in terms of continuous sets in this work as it has no practical relevance for the discussed modelling and simulation methods. We will henceforth consider any continuous set to be a real interval.

As continuous sets lead to difficulties for implementation sometimes a so-called **discretisation** is applied.

Definition 1.3: Discretisation

For a function $f : A \to \Gamma$ on a continuous set A a function

$$\hat{f} : \hat{A} \to \Gamma$$

that approximates f or $f|_{\hat{A}}$ is called a discretised version of f. The process of calculating $\hat{f}$ and $\hat{A}$ based on f and A is called discretisation or discretisation process. The difference between f and $\hat{f}$ is called discretisation error.

In most cases a discretisation has to be applied in order to execute a stated model on a computer. Hence, applying discretisation has usually a negative connotation as it is an unavoidable source for errors.

As it is going to be necessary to deal with both, discrete and continuous spaces, it is convenient to define two standard measure spaces which are used for this analysis.

- For discrete spaces Ω we usually consider $(\Omega, 2^\Omega, \mu)$ wherein 2^Ω stands for the **power set** and $\mu(A) := |A|$ denotes the **counting measure** on Ω.
- For continuous Banach spaces Ω the power set usually turns out to be unnecessarily large. Hence, we usually use $(\Omega, \mathcal{B}(\Omega), \mu)$ wherein $\mathcal{B}(\Omega)$ denotes the Borel σ-algebra or, more conveniently, **Borel sets** on Ω and μ the uniquely (w.r. to a normalisation factor) induced measure – the **Borel measure**.

The counting measure transforms any Lebuesgue/Bochner integral into a possibly infinite sum. Hence, calculus with functions like so-called random-variables on discrete spaces is usually easier than on continuous spaces.

We conclude this introducing section by briefly introducing a measure that is applicable for both, discrete and continuous sets. The so-called Dirac measure plays a key role in the mean-field applications in Chapter 4 as it is required in proofs.

Definition 1.4: Dirac Measure

Given a space Ω, a sigma algebra $A_\sigma(\Omega)$ and a fixed element $x \in \Omega$. The Dirac measure δ_x is defined as

$$\delta_x(A) = \begin{cases} 1, & x \in A \\ 0, & \text{else.} \end{cases} \tag{3.1}$$

With a real- or Banach- space valued function $f : \Omega \to \Gamma$ clearly

$$\int_\Omega f(y) d\delta_x(y) = f(x). \tag{3.2}$$

3.1.2 Random Variables

As the concept of random variables is vitally important for this work we will discuss it in more detail.

In general the sample space Ω of a probability space $(\Omega, A_\sigma(\Omega), P)$ is of a rather abstract nature. Although we are able to unite and cut sets we are (in general) not able to sum, subtract or perform other mathematical operations with its elements, which might yet be important to find specific distribution parameters. This becomes clear when thinking of the outcomes of a simple coin-toss experiment wherein $\Omega = \{\text{heads}, \text{tails}\}$ does not provide any structure to perform algebraic operations.

In order to solve this problem the concept of random-variables is introduced.

Definition 1.5: Random Variable

Given a probability space $(\Omega, A_\sigma(\Omega), P)$ and a measure space $(\Gamma, A_\sigma(\Gamma), \mu)$.
A $(A_\sigma(\Omega), A_\sigma(\Gamma))$ measurable function

$$X : \Omega \to \Gamma, \omega \mapsto X(\omega) \tag{3.3}$$

is called a Γ valued random variable on Ω.

Usually we expect Γ to have some metric vector-space structure. We will only consider real valued random numbers in this Chapter, but they can straight forwardly be extended to Banach space valued random-numbers.

For some cases it might be useful to define $\Gamma = \Omega$ with a different measure. An example for this would be the distribution of body weights among the population. While having some Gaussian probability measure P on the sample space $\mathbb{R}$ we could use the standard Lebesgue measure on $\mathbb{R}$ as the value-space.

The concept of random variables allows us to transfer the investigation of likelihood to a different space. Therefore, we define the image or push-forward measure of P.

Definition 1.6: Image Measure of P w.r. to X

Let X be a Γ valued random variable on Ω and $A \in A_\sigma(\Gamma)$ then

$$P^X(A) := P(X \in A) := P(X^{-1}(A)) \tag{3.4}$$

is called the image- or push-forward- measure of P w.r. to random variable X.

The inversion X^{-1} is naturally defined as

$$X^{-1}(A) := \{\omega \in \Omega : X(\omega) \in A\}.$$

The image measure transfers a stochastic problem from the sample space Ω to Γ as it is much more convenient for analysis. The random variable X is very often chosen as a bijective mapping

so that the underlying space does not need to be regarded at all – it is simply identified with the target space of the random variable. Naturally we write $P(X = c) := P(X \in \{c\})$. Clearly, any measurable function $f(X) : \Gamma \to \Gamma_2$ becomes a random variable $\Omega \to \Gamma_2$ itself.

As a direct consequence of the definition of the Lebuesge integral the image measure fulfils the transformation formula

$$\int_\Omega f(\omega)dP(\omega) = \int_{X^{-1}(\Gamma)} f(\omega)dP(\omega) = \int_\Gamma f(X)dP^X(X) \tag{3.5}$$

for measurable and integrable $f : \Gamma \to \Gamma$ which is probably the basis for all probability calculus with random variables and, therefore, also random processes.

The most obvious way to imagine this, is given by considering the so-called indicator function as a random variable.

Definition 1.7: Indicator Function

For arbitrary measure space Ω and $A \in A_\sigma(\Omega)$

$$\mathbb{1} : \Omega \to \{0, 1\}, x \mapsto \mathbb{1}_A(x) = \begin{cases} 1, x \in A \\ 0, x \notin A \end{cases}.$$

is called indicator function of A.

Clearly

$$P^{\mathbb{1}_A}(\{1\}) = P(\mathbb{1}_A = 1) = P(\mathbb{1}_A^{-1}(1)) = P(A)$$

and

$$P^{\mathbb{1}_A}(\{1\}) = \int_{\{1\}} dP^{\mathbb{1}_A}(X) = \int_{\mathbb{1}_A^{-1}(1)} dP(\omega) = \int_A dP(\omega) = P(A).$$

As the indicator function maps any space onto the finite state-space $\{0, 1\}$ it poses the basis for many discretisation concepts.

3.1.3 Density Functions

For practical applications a probability measure P is usually very difficult to define directly as the underlying sample space Ω is hardly known and difficult to describe. More conveniently the measure is implicitly defined using so-called density functions. This concept is motivated as follows:

Investigate a random variable $X : \Omega \to \Gamma$ and a measurable function $f : \Gamma \to \Gamma$. According to equation (3.5) the inconvenient Lebuesgue integral on Ω is already expressed by a Lebuesgue integral on Γ which is usually more handy. Yet it remains to find a transformation from $dP^X(X)$ to $d\mu(x)$, the defined measure on the measure space $(\Gamma, A_\sigma(\Gamma), \mu)$. For this purpose the following version of the Theorem of Radon-Nikodym can often be applied.

Theorem 1.1: Radon-Nikodym (Special Case)

Let μ be a σ-finite measure on Γ and $\forall A \in A_\sigma(\Gamma):\ \mu(A) = 0 \Rightarrow P^X(A) = 0$ (we say P^X is absolutely continuous w.r. to μ and write $P^X \ll \mu$) then there exists a μ integrable function g classically called **density** or **Radon-Nikodym density** so that

$$P^X(A) = \int_A g(x)d\mu(x) \tag{3.6}$$

and

$$\int_B f(X)dP^X = \int_B f(x)g(x)d\mu(x), \tag{3.7}$$

for all P^X integrable functions $f : \Gamma \to \Gamma$.

While the sigma-finiteness of a measure on a Banach space is usually given for practical applications, the absolute continuity of the measure is sometimes violated. A very prominent example for this is given by investigating the life expectancy of light bulbs as a real valued random number. While the one-element set $\{0\}$ has Borel measure $\mu(\{0\}) = 0$ yet the probability that the light bulb does not work at all, i.e. it survives 0 seconds, $P(X \in \{0\})$, is usually positive.

Hence, a density function is a handy feature of a distribution but its existence is not guaranteed.

Remark 1.2:

We will often write dP instead of dP^X if the random variable(s) which is/are used for the image measure will become clear, given the context or the integration variable, respectively.

3.1.4 Mean, Variance and Covariance

In order to investigate properties of a distribution we are interested in meaningful (deterministic) parameters that can be used to describe its behaviour. Surely the density function defined in Theorem 1.1 poses for such a parameter, but there are much quicker ways to get a basic idea of the behaviour of a random number. Moreover, we have already seen that it does not always exist.

The most prominent are mean or **expected value**, **variance**, and the **standard deviation**, which all belong to the so-called **moments** of a distribution.

Definition 1.8: Expected Value

Given a Γ valued random variable X, the scalar

$$\mu_X := \mathbb{E}(X) := \int_\Gamma X dP(X) \tag{3.8}$$

(if it exists) is denoted as expected value of X.

The expected value can be shown to be the most accurate guess for predicting the value of the random number. In case we are interested in the average of a measurable function $f(X)$ instead of X we may apply the following corollary.

Corollary 1.1: Law of the Unconscious Statistician

Given a Γ valued random variable X and a measurable function $f : \Gamma \to \Gamma$ then

$$\mathbb{E}(f(X)) = \int_\Gamma f(X) dP(X). \tag{3.9}$$

The irritating name of above corollary results from the problem, that the equation above is not treated as a law/theorem but, falsely, as the definition of the expected value itself as the statement seems so obvious. In that case the "statistician" was "unconscious" about the fact that $\mathbb{E}(f(X))$ is basically defined by the distribution of $f(X)$ and that it requires a simple transformation law to rewrite it using the distribution of X instead – therefore "Law of the Unconscious Statistician".

Definition 1.9: Variance and Standard Deviation

Given a Γ valued random variable X then the scalar

$$\sigma_X^2 := \mathbb{V}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) := \int_\Gamma (X - \mu_X)^2 dP(X) \tag{3.10}$$

(in case it exists) is denoted as variance of X. The square root σ of this positive number is denoted as standard deviation of X.

Given a second random variable $Y : \Omega \to \Gamma$ based on the same probability space, we can investigate the common distribution of X and Y by investigating the image measure on $\Gamma \times \Gamma$ via $= P^{X,Y}(A, B) = P(X \in A \wedge Y \in B) = P(X^{-1}(A) \cap Y^{-1}(B))$. This idea can be generalised as follows.

Definition 1.10: Common Distribution of Random Variables

Given Γ valued random variables $X_1, \ldots, X_n$ and $A_1, \ldots, A_n \in A_\sigma(\Gamma)$ then

$$P^{X_1,\ldots,X_n}((A_1,\ldots,A_n)) := P\left(X_1^{-1}(A_1) \cap X_2^{-1}(A_2) \cap \cdots \cap X_n^{-1}(A_n)\right) \tag{3.11}$$

is denoted as common image measure of the random variables. Hereby

$$P^{X_1,\ldots,X_n} : \bigtimes_{i=1}^{n} \Gamma \to [0,1].$$

We usually write

$$P(X_1 \in A_1 \wedge X_2 \in A_2 \wedge \cdots \wedge X_n \in A_n) := P^{X_1,\ldots,X_n}((A_1,\ldots,A_n)) \tag{3.12}$$

as it is easier to interpret.

Using this notation, the covariance of two random variables can be defined.

Definition 1.11: Covariance

Let P define the common image measure of X and Y then

$$\mathrm{Cov}(X,Y) := \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) = \int_{\Gamma\times\Gamma} (X - \mu_X)(Y - \mu_Y) dP(X,Y)$$

denotes the random variables' covariance.

Calculation with expected values and variances are usually done using their handy properties. We summarise them without proof in the following corollaries.

Corollary 1.2: Properties of Expected Values

The expected values of random numbers X, Y on $\mathbb{R}$ fulfil

- $\mathbb{E}(c) = c$ for all constants,
- $\mathbb{E}(cX + Y) = c\mathbb{E}(X) + \mathbb{E}(Y)$ (linearity), and
- $X \geq Y, a.s. \Rightarrow \mathbb{E}(X) \geq \mathbb{E}(Y)$ (monotonicity).

Corollary 1.3: Properties of Variances and Covariances

Variance and covariance of random numbers X, Y, Z on $\mathbb{R}$ fulfil

- $\mathbb{V}(X) = \text{Cov}(X, X)$,
- $\text{Cov}(c, X) = \text{Cov}(X, c) = 0$ for all constants c,
- $\text{Cov}(cX, Y) = c\text{Cov}(X, Y) = \text{Cov}(X, cY)$ for all constants c (multilinearity),
- $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$ (multilinearity),
- $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$ (multilinearity),
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ (symmetry),
- $\mathbb{V}(cX + d) = c^2\mathbb{V}(X)$ for all constants,
- $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ (Steiner Theorem) which results in

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

- $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2\text{Cov}(X, Y)$, and
- $\text{Cov}(X, Y)^2 \leq \mathbb{V}(X)\mathbb{V}(Y)$ (Cauchy-Schwartz inequality).

3.1.5 Conditional Probabilities and Bayesian Theorem

Investigating the execution of an experiment with two outputs, X and Y, very likely the knowledge of one of these output values has an influence on the probability of second one. Imagine rolling a six sided dice twice and define Y as the result of the first throw and X as the sum of both throws. In this experiment the knowledge of $Y = 1$ already indicates that $X > 7$ is impossible. The probabilistic term for this feature is **dependence** of two random variables which is formally defined as the opposite of **independence**.

Definition 1.12: Independent Random Variables

Two random variables $X, Y : \Omega \to \Gamma$ are called independent if $\forall A, B \in A_\sigma(\Gamma)$

$$P^{X,Y}((A, B)) = P(X \in A \wedge Y \in B) = P(X \in A)P(Y \in B) = P^X(A)P^Y(B). \tag{3.13}$$

Note, that this is equivalent to

$$\forall A, B \in A_\sigma(\Gamma): \quad P(X^{-1}(A) \cap Y^{-1}(B)) = P(X^{-1}(A))P(Y^{-1}(B)). \tag{3.14}$$

Although it seems to be more convenient to work with independent random variables, as the common probability image measure $P^{X,Y}$ can be written as a product of the individual measures P^X and P^Y, in reality dependence of two random events is a highly welcome feature: As mentioned, the knowledge of the value of either of the two random variables increases the chances of successfully predicting the value of the other. This is done by so-called conditional probability.

Definition 1.13: Conditional Probability of Random Variables (1)

For two random variables $X, Y : \Omega \to \Gamma$ and $P^Y(B) > 0$ the measure $P^{X|Y\in B}(A)$ defined via

$$P^{X|Y\in B}(A) := \frac{P^{X,Y}((A,B))}{P^Y(B)} = \frac{P(X^{-1}(A)\cap X^{-1}(B))}{P(Y^{-1}(B))}$$

is called conditional image measure of X dependent on $Y \in B$. We more conveniently write

$$P(X \in A|Y \in B) := P^{X|Y\in B}(A).$$

This definition, known as the **Formula of Bayes**, may looks harmless, but leads to a couple of problems regarding $P(Y \in B) = 0$. Say we investigated the conditional probability with respect to the one-point set $\{b\}$ for a real valued random number Y very likely $P(Y \in \{b\}) = 0$, which makes it impossible to define the conditional probability via Bayes's formula. Yet, investigation of the probability $P(X \in A|Y = b)$ definitely makes sense for phenomenological reasons.

Trivially deducing a definition for it from the conditional image measure $P(X \in A|Y \in B)$ by applying some limit $U_\epsilon(b) \underset{\epsilon\to 0}{\longrightarrow} \{b\}$ does not work either as Kolmogorow and Borel already showed in their paradoxon [Kolmogorov, A, 1931,Jaynes and Bretthorst, 2003] that the resulting limit is not unique. Hence, the conditional probability $P(X \in A|Y = b)$ cannot be defined with measure theoretic means alone, but poses a new concept.

Definition 1.14: Conditional Probability of Random Variables (2)

Define

$$P(X \in A|Y = y) := P(X(\omega) \in A|Y(\omega) = y) := P(X \in A \text{ under condition that } Y \in \{y\})$$

which denotes the conditional probability of $X \in A$ knowing that $Y = y$. Let $\mathcal{P}(A_\sigma(\Omega))$ stand for the set of probability measures on Γ it follows that

- $P : \Gamma \to \mathcal{P}(A_\sigma(\Gamma)) : y \mapsto P(X \in \cdot|Y = y)$ is a probability-measure- valued measurable function on Γ and
- $P : \Omega \to \mathcal{P}(A_\sigma(\Gamma)) : \omega \mapsto P(X \in \cdot|Y = Y(\omega)) =: P(X \in \cdot|Y)$ is a probability-measure- valued random variable on Ω.

For $P(Y \in B) > 0$ both definitions 1.13 and 1.14 are reunited again as

$$\int_B P(X \in A|Y)dP(Y) = P(X \in A|Y \in B)$$

can be proven. This implies the **Theorem of Total Probability**.

Corollary 1.4: Theorem of Total Probability

For two random variables X and Y on Γ and any $A \in A_\sigma(\Gamma)$

$$\mathbb{E}(P(X \in A|Y)) = \int_\Gamma P(X \in A|Y)dP(Y) = P(X \in A). \tag{3.15}$$

3.1.6 Conditional Expectation and Covariance

Similar to classic random variables also expected values w.r. to conditional probabilities can be calculated. Hereby we denote the expected valued of a random variable X or $f(X)$ with respect to the conditional measure $P^{X|Y}$, which itself is a random variable. Hence, also the resulting expected value is a random variable.

Definition 1.15: Conditional Expectation

The **random variable** $\mathbb{E}(X|Y)$ defined by

$$\mathbb{E}(X|Y) := \int_\Gamma X dP(X|Y) := \int_\Gamma X dP^{X|Y}(X) \tag{3.16}$$

is called conditional expectation of X w.r. to Y. The function

$$y \mapsto \mathbb{E}(X|Y = y) = \int_\Gamma X dP(X|Y = y)$$

is measurable and denotes the conditional expected value of X given $Y = y$.

Analogous to the Law of the Unconscious Statistician 1.1 it follows that

$$\mathbb{E}(f(X)|Y) = \int_\Gamma f(X)dP^{X|Y}(X).$$

Analogously conditional variance and covariance can be defined.

Definition 1.16: Conditional Variance and Covariance

The random variable $\mathbb{V}(X|Y)$ defined as

$$\mathbb{V}(X|Y) := \mathbb{E}((X - \mathbb{E}(X|Y))^2|Y) \tag{3.17}$$

is called conditional variance of X w.r. to Y. Moreover, random variable

$$\text{Cov}(X, Y|Z) := \mathbb{E}((X - \mathbb{E}(X|Z))(Y - \mathbb{E}(Y|Z))|Z) \tag{3.18}$$

is called conditional covariance of X and Y w.r. to Z. For specific $Y = y$ both are measurable functions.

The conditional expectation can be interpreted as the most feasible prognosis for the value of random variable X when knowing random variable Y. For the dice rolling example introduced earlier the conditional expectation for the sum of both dices X when knowing the result of the first dice Y would be $\mathbb{E}(X|Y) = Y + 3.5$ and can be seen to be a random variable. Similarly a measurable function $f(y) := \mathbb{E}(X|Y = y) = y + 3.5$ is defined. For a specific first throw result $Y = 2$ we get $f(2) = \mathbb{E}(X|Y = 2) = 2 + 3.5 = 5.5$.

Clearly, $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(Y + 3.5) = 3.5 + 3.5 = 7 = \mathbb{E}(X)$ holds, which can be proven to be a general law. We summarise this and other properties of mentioned conditional moments.

Corollary 1.5: Properties of Conditional Moments

In addition to all properties presented in Corollaries 1.2 and 1.3 conditional moments satisfy the following formulas.

- If X is independent from Y then $\mathbb{E}(X|Y) = \mathbb{E}(X)$ and $\mathbb{V}(X|Y) = \mathbb{V}(X)$ (Y does not provide any information)
- $\mathbb{E}(X|X) = X, \mathbb{V}(X|X) = 0$ and $\text{Cov}(X, Y|X) = 0$ (X is perfectly known)
- $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$ (Theorem of Total Expectation)
- $\mathbb{V}(X) = \mathbb{E}(\mathbb{V}(X|Y)) + \mathbb{V}(\mathbb{E}(X|Y))$ (Theorem of Total Variance)
- $$\text{Cov}(X, Y) = \mathbb{E}(\text{Cov}(X, Y|Z)) + \text{Cov}(\mathbb{E}(X, Y|Z)) \tag{3.19}$$
  (Theorem of Total Covariance)

3.2 Statistical Basics

In this work we will mostly use methods from so-called **descriptive statistics** and **mathematical statistics** which include estimation of moments and probability- or density- functions. **Exploratory statistics** mainly deals with hypothesis testing and seems to be of lesser use for stochastic analysis of simulation models.

Given a set of N drawn observations $x_1, x_2, \ldots, x_N$, henceforth called a **sample**, of independent, identically distributed (**iid**) random variables $X_1, \ldots, X_N$ with $X_i \sim D$, $i \in \{1, \ldots, N\}$, we aim to make most likely assumptions for moments of D or the distribution of D itself. In

general, any measurable function f that operates on a sample is called an **estimator** or estimation function if it estimates the value of an unknown parameter p of the common distribution $D = D(p)$ of $X_1, \dots, X_N$ via

$$f(x_1, \dots, x_N) \approx p.$$

Function $f(X_1, \dots, X_N)$ is a random number of which we usually expect to converge towards the constant parameter p (at least) in probability

$$f(X_1, \dots, X_N) \xrightarrow[N\to\infty]{p} p.$$

In that case we speak of a **consistent** estimator. If $\mathbb{E}(f(X_1, \dots, X_N)) = p$ the estimator is called **unbiased**.

3.2.1 Standard Estimators for Moments

The most famous estimation functions are clearly the estimators for moments.

Definition 2.1: Sample Mean

The estimator

$$\overline{X} := f(X_1, \dots, X_N) := \frac{1}{N} \sum_{i=1}^{N} X_i$$

is called sample mean. The resulting parameter

$$\overline{x} := f(x_1, \dots, x_N) = \frac{1}{N} \sum_{i=1}^{N} x_i$$

is a consistent and unbiased estimator for $\mathbb{E}(X_1) = \dots = \mathbb{E}(X_N)$.

While it is easy to see that the estimator is unbiased (linearity of the expected value) it is a little bit more difficult to show that it is consistent. The latter follows from the Law of Large Numbers (see later in Theorem 3.1). As a direct consequence of the latter also

$$\overline{f(X)} := \frac{1}{N} \sum_{i=1}^{N} f(X_i) \tag{3.20}$$

is a consistent and unbiased estimator for $\mathbb{E}(f(X_i))$ for every measurable function f. In case f additionally depends on other estimators $p_1, p_2, \dots$ on the sample set $X_1, \dots, X_N$ via $f(X_i, p_1, p_2, \dots)$ the resulting estimator is consistent as long as the individual estimators are consistent. Yet independent of the quality of the individual estimators $\overline{f(X, p_1, p_2, \dots)}$, it must not be unbiased anymore. The most prominent example for this is given by

$$\overline{(X - \overline{X})^2} = \frac{1}{N} \sum_{i=1}^{N} (X_i - \overline{X})^2$$

used as an estimator for the variance. It is consistent but can easily be calculated to violate the condition for unbiased estimation

$$\mathbb{E}\left(\overline{(X-\overline{X})^2}\right) \neq \mathbb{V}(X_i).$$

Yet there is a way to remove this bias by slight modification of the pre-set factor.

Definition 2.2: Sample Variance (unbiased)

The estimator function

$$f(X_1,\ldots,X_N) = \frac{1}{N-1}\sum_{i=1}^{N}(X_i-\overline{X})^2$$

is called sample variance. The resulting parameter

$$s^2 := f(x_1,\ldots,x_N) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\overline{x})^2$$

is a consistent and unbiased estimator for $\mathbb{V}(X_1) = \cdots = \mathbb{V}(X_N)$.

Clearly,

$$s := \sqrt{s^2} \tag{3.21}$$

is a consistent estimator for the standard deviation – the so-called sample standard deviation – but it is biased. Moreover, it can be proven that it is impossible to give a closed formula for an unbiased estimator for the standard deviation without additional knowledge of the underlying distribution (more about this problem is found in [Bolch, 1968]).

3.2.2 Estimators for Probability- and Density- Functions

While above estimators focus on finding approximation for scalar distribution parameters, it is possible to find estimators for the whole probability- or density- function as well. In this case the estimated parameter is not a scalar but can be interpreted as an executable function uniquely describing a distribution $\hat{P}$. It is determined so that the difference between $\hat{P}$ and the original unknown distribution P, is as small as possible. Most commonly this is done by estimating the distribution's density function f either by a histogram or by so-called kernel density estimation.

Definition 2.3: Kernel Function

A positive function $K : \mathbb{R} \to \mathbb{R}^+$ that integrates to one is called kernel function. Kernels are usually chosen symmetric and with $K(x) \leq K(0)$.

Definition 2.4: Kernel Density Estimation

Given N iid samples $x_1, \ldots, x_N$ then

$$\hat{f}(x) := \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) \tag{3.22}$$

approximates the distribution's density function if K is a kernel function and the bandwidth h is chosen reasonably small.

As it is not important for our applications we will not go into details about convergence speed or similar technical issues. More importantly a suitable kernel and a fitting bandwidth has to be chosen for its application. We want to introduce two commonly used kernels.

Definition 2.5: Gaussian Kernel

$$K(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{3.23}$$

Definition 2.6: Uniform Kernel

$$K(x) := \mathbb{1}_{\left[-\frac{1}{2}, \frac{1}{2}\right]}(x). \tag{3.24}$$

Clearly, the Gaussian kernel is highly suited for distributions which are close to a normal distribution but quite expensive with respect to computational resources. The uniform kernel evaluates very quickly but (usually) results in unsteady solutions.

The bandwidth h needs to be chosen as small as required, but not smaller. If it is too small, the resulting density estimation receives unrealistic fluctuations. However, if it is too large, the curve is too much flattened. For the Gaussian kernel the rule of thumb estimate (see [Silverman, 1998])

$$h \approx 1.06 \sqrt{s^2} \frac{1}{N^{\frac{1}{5}}} \tag{3.25}$$

can be proven to result in good approximating density estimates. Herein s^2 denotes the sample variance. It also turns out to be quite applicable for the uniform kernel.

Clearly, a probability density can be approximated with a standard (normed) histogram as well. Here the success of the density approximation depends on the number of bins which can be seen to pose for some analogue of the bandwidth in kernel density estimators. The difference between histogram and kernel density estimation is seen in Figure 3.1.

In case of solely discrete random variables, i.e. $\Gamma \cong A \subseteq \mathbb{N}$, the density function, which is usually called probability mass function in that case, can also be approximated using both

methods. Yet both concepts, the histogram and the kernel density estimation, lead to the same result. Note, that the only useful kernel for that application is given by

$$K(x) := \mathbb{1}_{\{0\}}(x).$$

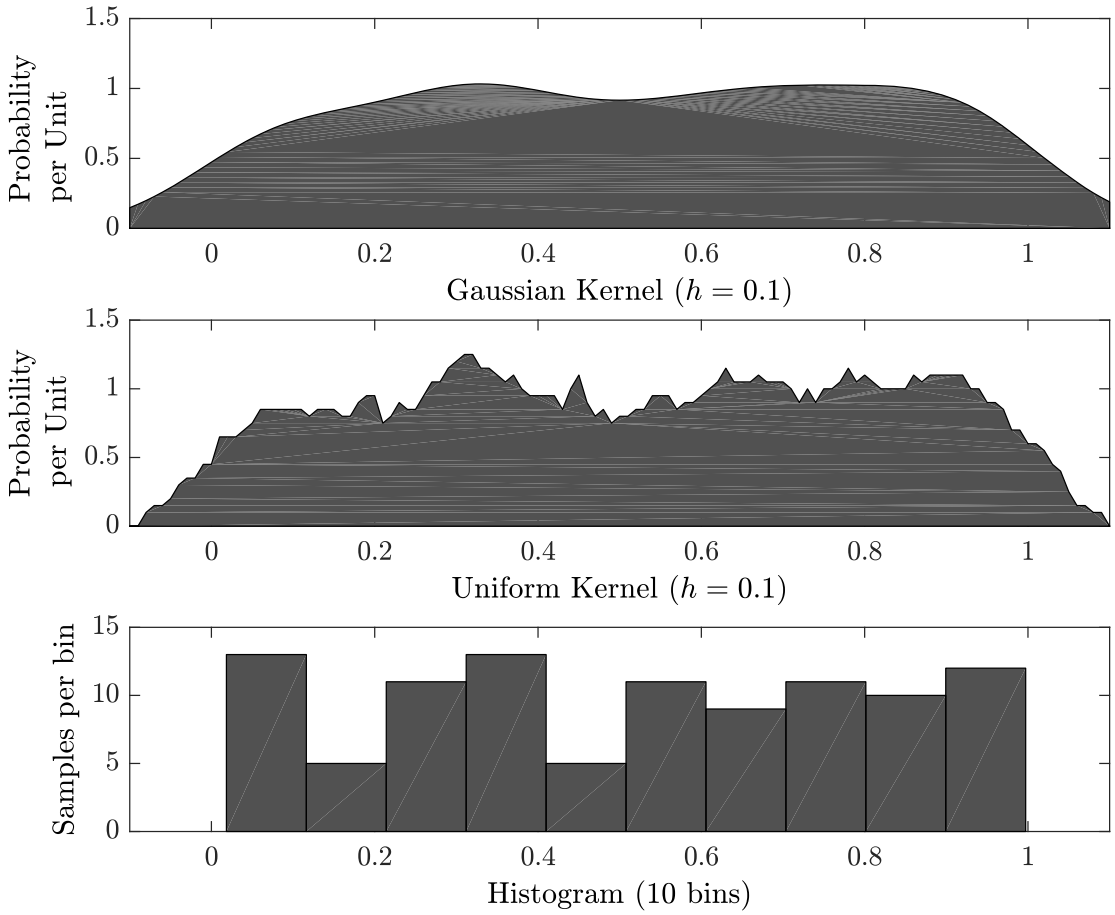


Figure 3.1: Comparison of different methods for density estimation with 100 uniformly $U(0,1)$ generated pseudo random numbers. While the upper two images show standard kernel density estimations with Gaussian and uniform kernel the third plot shows a standard histogram with a comparable number of bins.

3.3 Limit Theorems of Stochastic Sums

While probability theory deals with the question how knowledge about distributions can be used to make statements about experimental results, the field of statistics deals with the question how

experimental results can be used to make statements about underlying distributions. Surely both sciences strongly overlap and are usually applied together. A perfect example for this is the analysis of the distribution of stochastic sums, i.e. a sum

$$Y_N := \sum_{i=1}^{N} X_i^0 := \sum_{i=1}^{N} \frac{X_i - \mathbb{E}(X_i)}{\sqrt{\mathbb{V}(X_i)}} \tag{3.26}$$

of N random numbers X_i with finite expected value and variance. This case study is, of course, motivated by the estimation of moments like mean and variance which is usually done via summation. Nevertheless the analysis of the underlying distribution of Y_N based on the individual distributions of X_i is done using methods from probability theory. Moreover, these sums are playing a key role in the analysis of microscopic models via mean-field analysis.

When investigating stochastic sums three key-theorems can be seen to be vitally important. All three deal with the question what happens with the limit

$$Y_\infty^i := \lim_{N\to\infty} C_N \cdot \sum_{i=1}^{N} X_i^0 \tag{3.27}$$

for specific N-dependent constants C_N. Herein the following three cases will be investigateds:

1. The limit is a constant random number (Law of Large Numbers).
2. The limit is a random number that spreads (somehow) on a finite interval (Law of Iterated Logarithm).
3. The limit is a random number on $[-\infty, \infty]$ with finite variance (Central Limit Theorem).

3.3.1 Law of Large Numbers

The Law of Large Numbers considers $C_N := 1/N$ leading to

$$Y_N = \frac{1}{N} \sum_{i=1}^{N} X_i^0 = \overline{X^0}. \tag{3.28}$$

As $\overline{X^0}$ is an estimator for the mean of the individual random numbers X_i^0 we expect that

$$Y_N \to 0$$

as all $X_i^0 = \frac{X_i - \mathbb{E}(X_i)}{\sqrt{\mathbb{V}(X_i)}}$ are normed to have mean 0 and variance 1. If the random numbers X_i are independent this assumption is true, which is guaranteed by the Law of Large Numbers. Hereby two (main) versions exist, a strong and a weak one. They differ by the type of convergence.

Theorem 3.1: Law of Large Numbers (Kolmogorov)

If X_i are independent random numbers with finite variance and expected value then the stochastic sum $\overline{X^0}$ fulfils the Strong Law of Large Numbers

$$\lim_{N\to\infty} P(\overline{X^0} = 0) = 1 \quad \left(\Leftrightarrow \quad \overline{X^0} \xrightarrow{a.s.} 0\right), \tag{3.29}$$

which implies the Weak Law of Large Numbers

$$\forall \epsilon > 0 : \lim_{N\to\infty} P(\overline{X^0} > \epsilon) = 0 \quad \left(\Leftrightarrow \quad \overline{X^0} \xrightarrow{p} 0\right). \tag{3.30}$$

Clearly, for any $C_N < \frac{1}{N}$ we also obtain $Y_N \xrightarrow{a.s.} 0$ but $\frac{1}{N}$ plays a key role. In case all random numbers are identically distributed with finite mean μ and variance σ^2 we get

$$Y_N = \frac{1}{N}\sum_{i=1}^{N} \frac{X_i - \mu}{\sigma} = \frac{1}{\sigma}\left(\overline{X} - \mu\right) \xrightarrow{a.s.} 0 \quad \Rightarrow \quad \overline{X} \xrightarrow{a.s.} \mu. \tag{3.31}$$

3.3.2 Law of the Iterated Logarithm

The Law of the Iterated Logarithm considers $C_N := \frac{1}{\sqrt{2N\ln(\ln(N))}}$ which results in

$$Y_N = \frac{1}{\sqrt{2N\ln(\ln(N))}} \sum_{i=1}^{N} X_i^0 \tag{3.32}$$

Although the shape of the limit distribution depends on the individual distributions of X_i the law of the iterated logarithm states that the support of Y_∞ is limited.

Theorem 3.2: Law or the Iterated Logarithm (Hartman & Windtner)

If X_i are independent random numbers with finite variance and expected value then

$$\limsup_{N\to\infty} \frac{1}{\sqrt{2N\ln(\ln(N))}} \sum_{i=1}^{N} X_i^0 = 1 \text{ a.s. and} \tag{3.33}$$

$$\liminf_{N\to\infty} \frac{1}{\sqrt{2N\ln(\ln(N))}} \sum_{i=1}^{N} X_i^0 = -1 \text{ a.s..} \tag{3.34}$$

For identically distributed random numbers with mean μ and variance σ^2 we observe that

$$\lim_{N\to\infty} \frac{N}{\sqrt{2\sigma^2 N\ln(\ln(N))}} \left|\overline{X} - \mu\right| = \lim_{N\to\infty} \frac{1}{\sqrt{2N\ln(\ln(N))}} \left|\sum_{i=1}^{N} X_i^0\right| = 1.$$

Thus, $\frac{\sqrt{2\sigma^2 N \ln(\ln(N))}}{N} = \sqrt{\frac{2\sigma^2 \ln(\ln(N))}{N}}$ can be used as an estimator for the error of the sample mean via

$$\left| \frac{1}{N} \sum_{i=1}^{N} X_i - \mu \right| = \mathcal{O}\left(\sqrt{\frac{\ln(\ln(N))}{N}} \right) \tag{3.35}$$

and for reasonable large N

$$\left| \frac{1}{N} \sum_{i=1}^{N} X_i - \mu \right| \leq \sqrt{\frac{2\sigma^2 \ln(\ln(N))}{N}} \approx \sqrt{\frac{2s^2 \ln(\ln(N))}{N}}. \tag{3.36}$$

3.3.3 Central Limit Theorem

Finally the Central Limit Theorem considers

$$C_N := \frac{1}{\sqrt{N}}$$

and results in a very prominent limit distribution for Y_∞ – the normal or Gaussian distribution.

Theorem 3.3: Central Limit Theorem (Lyapunov)

If X_i are independent random numbers with finite variance and expected value then

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i^0 = Y_\infty \sim \mathcal{N}(0,1), \tag{3.37}$$

in case the individual variances $\mathbb{V}(X_i)$ do "not grow too fast" with N – i.e. they fulfil the Ljapunov condition

$$\lim_{N\to\infty} \frac{1}{\mathbb{V}(X_N)^{\delta+2}} \sum_{i=1}^{N} \mathbb{E}(|X_i - \mathbb{E}(X_i)|^{2+\delta}) = 0 \tag{3.38}$$

for some $\delta > 0$.

The first limit needs to be understood in distribution.

The most important feature of this theorem is the fact that (in contrary to the Iterated-Logarithm Theorem 3.2) the limit distribution is independent from the individual distributions of X_i. A direct consequence of the Theorem is given for identically distributed random numbers X_i with $\mathbb{E}(X_i) = \mu$ and $\mathbb{V}(X_i) = \sigma^2$ by

$$\sqrt{N}\overline{X} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} X_i \xrightarrow{d} Y_\infty \sim \mathcal{N}(\mu, \sigma). \tag{3.39}$$

3.3.4 Extensions and Limitations of the Three Limit Theorems

Above limit theorems are of vital importance in descriptive statistics wherein the convergence of sums implies the convergence of specific estimators. In reality observed samples are very unlikely guaranteed to be independent from each other which leads to doubts about whether

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i^0 \underset{N\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

as all three limit theorems are basically defined for independent samples. Yet it might be important to know if especially $\mathbb{V}(\frac{1}{N}\sum_{i=1}^{N} X_i^0)$ disappears for $N \to \infty$ (which would be a direct consequence of all three theorems).

Rigorous research about sums of **dependent** random samples started in the middle of the last century. It is still a heavily researched area and it can be seen as a part of ergodicity theory. This research is justified as experiments clearly imply that the Central Limit Theorem (and consequently the Law of Large Numbers) also holds for certain series of dependent random numbers. Nevertheless there are very simple examples showing that this validity is not a general one. Define $X_1 \sim \delta_{0.5}(\{-1,1\})$ and

$$P(X_{i+1}=1|X_i=y)=\mathbb{1}_y(1), \quad P(X_{i+1}=-1|X_i=y)=\mathbb{1}_y(-1).$$

Clearly, $X_i = X_i^0$ and $\sum_{i=1}^{N} X_i = \sum_{i=1}^{N} X_i^0 \in \{-N, N\}$ which leads to

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i^0 \underset{N\to\infty}{\overset{a.s.}{\longrightarrow}} \pm\infty \tag{3.40}$$

instead of $\mathcal{N}(0,1)$.

It follows that a general extension to dependent random numbers is impossible. Most concepts that prove conservation of the validity of the Central Limit Theorem for dependent series use concepts that restrict the level of dependency between the random variables. To our knowledge Murray Rosenblatt [Rosenblatt, 1956] was the first who stated that the Central Limit Theorem 3.3 remains valid for sequences of random numbers which fulfil a so-called **strong mixing condition**. Using our notation a sequence of random numbers is called **strongly mixing** if

$$\forall A,B \in A_\sigma(\Gamma):\ P(X_i \in A \wedge X_{i+j} \in B) - P(X_i \in A)P(X_j \in B) \underset{|i-j|\to\infty}{\longrightarrow} 0. \tag{3.41}$$

Sloppily spoken, the dependence of X_i and X_j disappears when $|i-j|$ gets sufficiently large.

Applications for this Theorem are easily found when investigating random time series wherein

$$X_{i+1} = f(X_i)$$

for some randomness-containing function f that maps one random number onto the subsequent one. As the relation between X_i and X_j is given by $|i-j|$ times recursively applying f it is very likely that the random parts in f will dominate the influence of the input, in this case X_i,

for sufficiently large recursion depth. Note, that the given example which lead to (3.40) violated the strong mixing condition as f was completely deterministic and

$$\underbrace{P(X_i = 1|X_j = -1)}_{0} - \underbrace{P(X_i = 1)}_{\frac{1}{2}}\underbrace{P(X_j = -1)}_{\frac{1}{2}} \underset{|i-j|\to\infty}{\longrightarrow} -\frac{1}{4} \neq 0.$$

We can generalise the idea of a random time series by increasing the recursion order like

$$X_{i+1} = f(X_{i-m}, \ldots, X_i),\ m \in \mathbb{N},$$

or by interpreting the series as a spatial field via

$$X_i = f(X_{i-m}, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{i+m'}),\ m, m' \in \mathbb{N}. \tag{3.42}$$

which should conserve the validity of the Central Limit Theorem if f contains sufficient level of randomness. Based on this idea we are basically able to analyse microscopic models with respect to their limit distribution. Assuming that the sum of random numbers describes the aggregated numbers in a microscopic model, equation (3.42) is comparable with the concept of neighbourhood in a cellular automata (CA) seen in Section 2.1.2. Hence, stochastic CAs are very likely to fulfil a Gaussian limit distribution on the aggregate level if their transition function contained "sufficient randomness". For deterministic CAs for which randomness is only obtained due to a (possibly) random initial configuration this argument cannot be applied. This will be verified by a very prominent test-case in Section 5.4.

In agent-based models (Section 2.1.3) it is possible that agents might interact on a global level or via complex networks which directly violates the strong mixing condition. Consider an agent-based model wherein a designated group leader, agent 1, has a random opinion X_1 which influences not only its neighbours, but all agents in the complete environment. So problems with central limit conditions appear and the limit distribution of the aggregated numbers might have an unexpected shape. Some of these problems are discussed in [Kitzler, 2016].

In summary, limit theorems for stochastic sums are very useful for the analysis of estimators applied on observed independent random experiments. In case the experiments are executed serially, it is very likely that the strong mixing condition is fulfilled even for slightly dependent experimental results. In case they are executed in parallel, the condition is very likely to be violated if dependence appears on the global level. For more information about limits of sums of dependent random variables, the reader is referred to [Rio, 2013].

Unfortunately the three discussed Theorems can hardly be used to gain quantitative statements about microscopic models as, first, they are not constructive and, second, do not allow to analyse time and state as separate state variables of a model. I.e. either we may fix a specific point in time and approximate the limit for infinite number of sub-models or we discretise the time-space and investigate the limit distribution for $t \to \infty$, but we cannot combine these ideas. A more powerful concept needs to be introduced to solve this problem for formal mean-field analysis.

3.4 Stochastic Processes

One of the few common denominators of microscopic models is that their system components somehow change with time. Moreover, as most of them contain stochastic elements the concept of random- or, more frequently, **stochastic- processes** turn out to pose for a applicable mathematical foundation to formalise them.

A stochastic process is usually defined as a time (we will write T for the time-domain to the process) dependent random variable.

Definition 4.1: Stochastic Process

Let Ω be a probability space. Γ a measure space and $T \subseteq \mathbb{R}^+$ the so-called **time-space**. A function

$$X : \Omega \times T \to \Gamma : (\omega, t) \mapsto X(\omega, t)$$

is called stochastic process, if $\forall t \in T : X(\cdot, t)$ is $(A_\sigma(\Omega), A_\sigma(\Gamma))$-measurable. Space Γ is called **state-space** of the process.

Defining $\mathcal{M}(\Omega, \Gamma)$ as the set of measurable functions from Ω to Γ it is possible to investigate

$$X : T \to \mathcal{M}(\Omega, \Gamma) : t \mapsto X(\cdot, t) =: X(t)(\cdot)$$

as a random-variable valued function in time. It is henceforth called path or **trajectory** of the process. For fixed ω an evaluation

$$t \to X(t)(\omega)$$

is a measurable function in time and called **sample trajectory** or sample path. The collection of all possible trajectories $\{t \to X(t)(\omega), \omega \in \Omega\}$ is called **path-space** of the process.

We will henceforth mostly use the representation of the stochastic process in form of its trajectory.

On the one hand, the summary of all state variables in a microscopic model can clearly be formalised as one stochastic process with arbitrarily complex state-space while every simulation execution corresponds to a trajectory of the process. On the other hand the states of all individual sub-models in the microscopic model can, as well, be formalised as a (very likely dependent-) stochastic processes. Hence, the stochastic process *model* is uniquely defined as the summary (this is not trivial and may include other stochastic processes) of all individual stochastic processes *individually*. This feature makes stochastic processes to the probably most valuable concept for formalising and analysing microscopic models.

Above definition of a stochastic process is very general and does not contain any advise how to (uniquely) describe a process using mathematical terms. While a random variable can, for example, be defined via its probability density or probability mass function it is easy to see that this concept is not very practicable for stochastic processes. We would require to define one

density function for every point in the time-space T which is neither convenient nor picturesque as a system evolving in time.

Consequently, stochastic processes use a different definition-concept which is based on how the trajectory of the process is expected to evolve given the already sampled path. The **methods to define stochastic processes** via their trajectory focus on the following **classification** of stochastic processes.

It is generally useful to demand that the path of a stochastic process only depends on its prior states. This is phenomenologically founded thinking of the causality of *time* as the underlying space. Hence, the trajectory of a stochastic process is, in the general case, defined by the conditional probabilities

$$P(X(t) \in A | X(s) = x(s), s < t), \quad A \in A_\sigma(\Gamma), \forall s : x(s) \in \Gamma. \tag{3.43}$$

Analogous to other time evolution systems like differential equations, also an initial condition needs to be given which, in this case, poses for an initial distribution $\forall A : P(X(t_0) \in A)$.

For applications this concept is quickly seen to be too general to be applicable. It is, essentially, infinitely-dimensional as we would have to define probabilities for any element of the path space. Hence, a lot of classes are defined to simplify a unique definition of the process.

3.4.1 Classification with Respect to Space Structure

The classification with respect to input and output formats of the trajectory plays a key role to simplify the definition. Hereby we refer to the set-specific basic properties of the spaces T and Γ. We recall the definitions for discrete and continuous sets 1.1 and 1.2.

Definition 4.2: State Discrete/Continuous Stochastic Process

A stochastic process

$$X : \Omega \times T \to \Gamma : (\omega, t) \mapsto X(\omega, t)$$

is called state, value or spatially discrete/continuous if Γ is a discrete/continuous set.

As with classical random numbers, the distinction between discrete and continuous state-space has a major influence on how the random process can be defined and analysed. For state discrete processes the definition given in equation (3.43) can be shortened to

$$P(X(t) = x | X(s) = x(s), s < t), \quad x, x(s) \in \Gamma \tag{3.44}$$

which might even lead to a finite number of different probabilities. This can moreover be summarised in a matrix with row-dimension $|\Gamma|$. Moreover, trajectories of state discrete processes can be seen as piecewise constant, while trajectories of state continuous processes may result in arbitrary functions on T.

time \ state	discrete	continuous
discrete	The trajectory can be imagined as a vector or sequence on a time-space grid.	The trajectory is a vector or sequence with arbitrary values.
continuous	The trajectory is piecewise constant.	The trajectory may have an arbitrarily complex shape.

Table 3.1: Classification of stochastic processes with respect to different space formats. Impact of discrete and continuous time- and state- spaces on the trajectory.

Definition 4.3: Time-Discrete/Continuous Stochastic Process

A stochastic process

$$X : \Omega \times T \to \Gamma : (\omega, t) \mapsto X(\omega, t)$$

is called time-discrete/continuous if T is a discrete/continuous set.

For discrete time processes it is convenient to write the time-space as a vector $T = (t_0, \ldots, t_{\text{end}})^T$ or sequence $T = (t_i)_{i=1}^{\infty}$. Clearly, the discreteness of the input (time) leads to a discrete output (the trajectory). Hence, $X_i := X(t_i)$ can be seen as a vector or a sequence as well. A definition in form of equation (3.43) rewrites to

$$P(X(t_{i+1}) \in A | X(t_j) = x_j, j \in \{1 \ldots i\}), \quad A \in A_\sigma(\Gamma), \forall j : x_j \in \Gamma. \tag{3.45}$$

Time-continuous processes result in continuous output paths, which can be drawn as a closed (possibly unsteady) line in a time-state diagram.

We summarise the classification of stochastic processes via the influence of different time- and state- spaces on the trajectory in Table 3.1.

3.4.2 Classification with Respect to Memory

Considering equation (3.43) a stochastic process evolves with a probability depending on the past i.e. the former states of the process. Knowing the total path of the stochastic process until a certain point in time t the distribution for its future is uniquely defined. Let

$$T(t) := T \cap [0, t) \tag{3.46}$$

and define $m(t)$ as the supremum of the set

$$\{q \in [0, t] : \forall A, x(s) : \\ P(X(t) \in A | X(s) = x(s), s \in T(t)) = P(X(t) \in A | X(s) = x(s), s \in [t - q, t) \cap T)\}, \tag{3.47}$$

then $m(t)$ is called the memory of the process as the process needs to "remember" the last $m(t)$ time instances of it's past to uniquely define its future distribution. Only if $m(t) = t$ the process requires its entire past to evolve uniquely.

It is possible to go one step further: In almost all cases it is possible to find a countable subset $D(t)$ of $T(t)$ that fulfils

$$P(X(t) \in A | X(s) = x(s), s \in T(t)) = P(X(t) \in A | X(s) = x(s), s \in D(t))$$

for all A and $x(s)$. Hence, it is possible to find a transition mapping that does not depend on an infinite dimensional function $x(s)$ but on a countable number of previous states of the process. We call this property **separability**.

Definition 4.4: Separable Process

A stochastic process $X : T \times \Omega \to \Gamma$ is called separable if for any $t \in [0, t)$ there is a countable set $D(t) \subset [0, t)$ so that

$$P(X(t) \in A | X(s) = x(s), s \in T(t)) = P(X(t) \in A | X(s) = x(s), s \in D(t)) \quad (3.48)$$

for all $A \in A_\sigma(\Gamma)$ and all possible trajectories $x(s) : [0, t] \to \Gamma$.

Consequently, there is a series of time instants t_i and states $X(t_i) = x_i$ so that

$$\begin{aligned} P(X(t) \in A | X(s) = x(s), s < t) \\ &= P(X(t) \in A | X(s) = x(s), s \in T(t)) \\ &= P(X(t) \in A | X(t_1) = x_1, X(t_2) = x_2, \dots). \end{aligned} \quad (3.49)$$

This definition of separability is unusual, but, in comparison with the classical one, we consider it for more picturesque. Mathematician Josef Doob proved that any stochastic process on $\mathbb{R}^n$ has a separable version. Hence, most stochastic processes used as models for real world applications are separable as well. We will henceforth only discuss separable processes.

Using the concept of separability it is possible to introduce another important feature of a stochastic process with respect to memory.

Definition 4.5: Memoryless

Given a separable stochastic process $X : T \times \Omega \to \Gamma$. If for any vector with arbitrary length $m \in \mathbb{N}$ of descending points $t > t_1 > t_2 > \dots > t_m$ in T, set $A \in A_\sigma(\Gamma)$ and values $x_1, \dots, x_m \in \Gamma$

$$P(X(t) \in A | X(t_1) = x_1, \dots, X(t_m) = x_m) = P(X(t) \in A | X(t_1) = x_1), \quad (3.50)$$

the process is called memoryless.

The conditional probability function of the process does not depend on any state of the process which took place before the last observation x_1. Therefore, the process does not "memorise"

anything which is not directly "observable" anymore. Hence, the term "memoryless" is justified. For time-continuous processes we receive $m(t) = 0$ and for time-discrete processes (with equidistant time-steps of length 1) $m(t) = 1$ using the notation of (3.47).

This condition is probably the most powerful restriction to a stochastic process and is frequently called **Markov condition**. Consequently, memoryless processes are usually called **Markov processes** which, in case they are moreover time-discrete, may be called **Markov chains**. The Markov condition implies a number of useful properties of which the following two are probably the most important.

First, all sample trajectories of memoryless processes are right-continuous. This means that $\forall A \in A_\sigma(\Gamma)$: the set $\{t \in T : x(t) \in A\}$ has a minimum or is empty. Hence, there is always a specific point in time when the process enters a specific state or a set of states.

Second, the process is uniquely defined by a mapping $\Gamma \times \Gamma \times T \to \mathbb{R}^+$ that indicates how likely the process switches from one state in Γ to any other state in Γ at a given time. A switch between different states is typically called **transition**. For discrete time processes this mapping directly relates to the so-called **transition probabilities**, i.e. the probability that the process switches from one state to a different state at time t. For continuous-time processes it is better to interpret the mapping as switching tendencies between states – we speak of so-called **transition rates**. In case the state-space is discrete, the mapping can be stated as a possibly infinite dimensional time-dependent matrix. More about this transition mapping is given in the next sections (Section 3.4.3 and 3.4.4).

Not completely memoryless stochastic processes, i.e. processes with small memory, are usually defined only for time-discrete processes. A time-discrete process that fulfils

$$\begin{aligned} \forall n > m: \quad & P(X(t_i) \in A | X(t_{i-1}) = x_{i-1}, \ldots, X(t_{i-m}) = x_{i-m}) = \\ & = P(X(t_i) \in A | X(t_{i-1}) = x_{i-1}, \ldots, X(t_{i-n}) = x_{i-n}) \end{aligned} \tag{3.51}$$

for all ascending points in time $t_{i-n} < \cdots < t_i$, set A, and values $x_1, \ldots, x_n$ is called Markov chain of m^{th} order. Note, that every n^{th} order Markov chain on Γ can be redefined as a standard Markov Chain X' on the extended state-space Γ^m. Defining

$$X'(t_i) := (X(t_i), X(t_{i-1}), \ldots, X(t_{i-m+1}))^T . \tag{3.52}$$

then X' follows a Markov chain of first order. This strategy can be imagined as "saving" the memory, i.e. the previous states of the process, as additional vector entries to the state. We give a short example for this:

Example 4.1: Fibonacci Sequence

Consider the recursive Fibonacci sequence $x_n = x_{n-1} + x_{n-2}$ interpreted as a discrete-time, discrete-space stochastic process via

$$P(X(t_i) \in A | X(t_{i-1}) = x_{i-1}, X(t_{i-2}) = x_{i-2}) = \begin{cases} 1, & x_{i-1} + x_{i-2} \in A \\ 0, & \text{otherwise.} \end{cases} \tag{3.53}$$

Clearly, a (deterministic) second order Markov chain on $\mathbb{R}$ is given as the probabilities for the next state not only depend on the current state x_{i-1}, but also on the state before. Defining $X'(t_i) := (X(t_i), X(t_{i-1}))^T$ we deduce

$$P\left(X'(t_i) \in A | X'(t_{i-1}) = (x_{i-1}, x_{i-2})^T\right) = \begin{cases} 1, & (x_{i-1} + x_{i-2}, x_{i-1})^T \in A \\ 0, & \text{otherwise.} \end{cases} \tag{3.54}$$

and receive a standard Markov chain on $\mathbb{R}^2$.

We will now take a closer look on mentioned transition mapping of stochastic processes which was already stated to be the most convenient way to define the dynamics of a stochastic process. Herein we briefly distinguish between time-discrete and time-continuous stochastic processes.

3.4.3 Transition Probabilities

Any process used in this and the next section can be considered as memoryless.

The term **transition** or, more precisely, state-transition is used to describe that a random process switches from one to a different state. I.e. for a stochastic process X with state-space Γ there is a certain point t_t in time so that $X(t_t^-) = x_1$ and $X(t_t^+) = x_2$ for two different elements $x_1, x_2 \in \Gamma$. As memoryless processes have right-continuous paths $X(t_t) = x_2$ moreover.

Defining a process via its transitions requires the definition of the so-called transition probabilities. Hereby we refer to the probability that a process being in state x_1 has a transition to a different state in a certain interval.

Definition 4.6: Transition Probability

Given a stochastic process X, two points in time $t > t_1 \in T$, a state x_1 and a set A then

$$P : \Gamma \times T \times A_\sigma(\Gamma) \times T \to [0, 1] : \ (x_1, t_1, A, t) \mapsto P(X(t) \in A | X(t_1) = x_1) \tag{3.55}$$

is called transition probability of X from x_1 to (a state inside) A. Analogous to the conditional probability measure

$$P(X(t) \in A | X(t_1))$$

is a measure-valued random variable.

This definition is crucial for the analysis methods of microscopic models in the next chapters. In case the transition probabilities are absolutely continuous (as a measure) with respect to the measure of Γ there exists a **transition-density**

$$\tilde{P} : \Gamma \times T \times \Gamma \times T \to \mathbb{R}^+ : \ (x_1, t_1, x_2, t) \mapsto \tilde{P}(X(t) = x_2 | X(t_1) = x_1) \tag{3.56}$$

with

$$\int_A \tilde{P}(X(t) = x | X(t_1) = x_1) d\mu(x) = P(X(t) \in A | X(t_1) = x_1). \tag{3.57}$$

Clearly, defining the transition probabilities for all points in time the dynamics of a memoryless process are uniquely defined. For so-called **homogeneous** processes the transition probability does not depend on two points in time, but only on the length of the time interval in between:

$$P(X(t) \in A | X(t_1) = x_1) = P(X(t - t_1) \in A | X(0) = x_1). \tag{3.58}$$

3.4.4 Transition Rates

The idea of defining the dynamics of a memoryless process via its transition probabilities

$$P(X(t + h) \in A | X(t) = a)$$

easily leads to enormous overhead as the transition probability needs to be defined not only for all times, a-prior and a-posterior states, but also for all time-differences. As a result of the Markov property and the Theorem or Total Probability (3.15)

$$\begin{aligned} &P(X(t + 2h) \in A | X(t) = x) \\ &\quad = \int_\Gamma P(X(t + 2h) \in A | X(t + h) = Y, X(t) = x) dP(Y | X(t) = x) \\ &\quad = \int_\Gamma P(X(t + 2h) \in A | X(t + h) = Y) dP(Y | X(t) = x) \end{aligned} \tag{3.59}$$

holds. Hence, transition probabilities for time-difference of $2h$ are uniquely defined by the transition probabilities for time-difference h and consequently by infinitesimal time-differences if this bisection process is repeated. Note, that equation (3.59) is called Equation of Chapman and Kolmogorov. As a result memoryless processes can be defined via differential expressions

$$\omega(t, x, A) := \lim_{h \to 0} \frac{P(X(t + h) \in A | X(t) = x) - P(X(t) \in A | X(t) = x)}{h},$$

as they uniquely define the dynamics of the process in infinitesimal steps.

To be precise:

Definition 4.7: Transition Rate

For fixed $x \in \Gamma$, $t \in T$ and $A \in A_\sigma(\Gamma)$ the limit (as long as it exists)

$$\omega(t,x,A) := \lim_{t_1 \to \inf\{t_2 \in T : t_2 > t\}} \frac{P(X(t_1) \in A | X(t) = x) - P(X(t) \in A | X(t) = x)}{t_1 - t} \tag{3.60}$$

is called **transition rate** (from x to A).

The unusual choice of the limit was necessary to regard time-discrete processes as well. Phenomenologically,

$$P(X(t) \in A | X(t) = x) = \mathbb{1}_A(x). \tag{3.61}$$

Hence, for all $x \notin A$ transitions rates are always positive while $x \in A$ leads to a negative transition rate.

Remark 4.1:

There are also definitions of transition rates that exclude sets A, that contain element x from their definition space to avoid irregularities (see later in the discussion about regularity of processes). This is legitimate as it is for practical applications. It not necessary to define the rate that a process remains in its state as it results from the rates that a process leaves it.

For time-discrete processes the transition rates always compute to

$$\omega(t_i, x, A) = \frac{P(X(t_{i+1}) \in A | X(t_i) = x) - \mathbb{1}_A(x)}{t_{i+1} - t_i}.$$

By rules of probability calculus $\omega(t_i, x, \sum_{i=1} B_i) = \sum_{i=1} \omega(t_i, x, B_i)$ can be shown for disjoint sets B_i which makes ω a signed measure. Moreover, $\omega(t, x, A) = -w(t, x, A^c)$ and hence

$$\int_\Gamma \omega(t_i, x, dy) = 0.$$

Spatially discrete processes with $\Gamma = \{x_1, \ldots, x_m\}$, ω can be summarised in a matrix

$$W(t) = (\omega(t, x_i, x_j))_{i,j=1}^m .$$

This matrix is strongly related but not equivalent with the standard transition matrix, which is only defined for spatially and temporally discrete Markov processes. According to the last observation rows of matrix $W(t)$ sum to 0 while rows and columns of the standard transition matrix

$$V(t_k) := (P(X(t_{k+1}) = x_i | X(t_k) = x_j))_{i,j=1}^m$$

sum to 1.

As implicitly mentioned in the definition of the transition rates 4.7 it is not always guaranteed that such a function mapping (to $\mathbb{R}$) really exists as it might occur that $\omega(t, x, A) = \infty$ for certain processes. Consider a process X with transition probabilities

$$\lim_{h\to 0} P(X(t+h) \in \{1\}|X(t) = 0) = 1, \quad \lim_{h\to 0} P(X(t+h) \in \{0\}|X(t) = 1) = 1.$$

Clearly, this process is memoryless and fully defined. Yet

$$\omega(t, 0, \{1\}) = \lim_{h\to 0} \frac{P(X(t+h) \in \{1\}|X(t) = 0)}{h} = \infty.$$

Reason for this irregularity is the problem that the process makes an infinite number of unsteady switches between 0 and 1 in any given finite time interval. Consequently, it is not a useful model to depict any given real system.

To avoid problems like this, which occurs for **continuous time processes only**, two restrictions are convenient. Sloppily spoken:

Definition 4.8: Regular time-continuous Markov process

If, during finite time intervals, the process trajectory almost surely has a finite number of **unsteady** state-transitions the process is denoted as **regular**.

This condition can be shown to be equivalent with

$$\lim_{h\to 0} \frac{P(X(t+h) \in \{1\}|X(t) = 0)}{h} = \omega(t, s, A) < \infty \tag{3.62}$$

for all compact A with $s \notin A$ and results in

$$P(X(t+h) \in A|X(t) = s) = h\omega(t, s, A) + \mathcal{O}(h^2). \tag{3.63}$$

We will henceforth call the latter **Poisson property**. For a proof of this statement the reader is referred to standard literature for Markov processes.

The highlighted adverb “unsteady” indicates that there are regular processes that allow an infinite number of transitions during a finite time interval as long as the transitions are continuous, i.e. with $h \to 0$ also $\|X(t+h) - X(t)\| \to 0$. According to this we distinguish between three types of processes.

Definition 4.9: Jump /Diffusion process

For a continuous metric space Γ with induced topology the following types of Markov processes can be distinguished:

- In case $\omega(t, s, A) = 0$ for all closed A with $s \notin A$ then the process is called (clean) **diffusion-process**.
- In case $\omega(t, s, A) < \infty$ for all A with $s \notin A$ then the process is called (clean) **jump process**.
- In case $\omega(t, s, A) < \infty$ for all closed A and there is at least one A with $s \notin A$ for which $\omega(t, s, A) = \infty$, then the process is called **jump-diffusion-process**.

Classically above distinction is only defined for $\Gamma = \mathbb{R}$, but an extension of this idea to arbitrary metric spaces does not pose a problem. **We will henceforth only consider regular jump processes in this work**.

Finally, we want to introduce a term that poses for the density pendent of the transition rates.

Definition 4.10: Transition Kernel

Let X be a regular time-continuous Markov process and

$$\{\omega(t, x, A),\ A \in A_\sigma(\Gamma), x \in \Gamma\}$$

be the family of all transition rates, then a function $\tilde{\omega}$ that fulfils

$$\int_A \tilde{\omega}(t, x, y)dy = \omega(t, x, A) \tag{3.64}$$

for all A is called transition kernel of the process. This concept is closely related to the transition density $\tilde{P}$ (see (3.56)) as

$$\tilde{\omega}(t, x, y) = \lim_{h \to \infty} \frac{\tilde{P}(X(t + h) = y | X(t) = x)}{h} \tag{3.65}$$

holds (as long as both exist). We write ω instead of $\tilde{\omega}$ if it is clear by context, that a kernel is used.

CHAPTER 4

Mean Field Analysis

In this chapter we want to introduce a method that can be applied to analyse a given and fully formalised microscopic model by a much simpler macroscopic model - a so-called **mean-field model**. This simplified model is usually found in form of ordinary or partial differential equations but may be found as difference equation as well. The concept of using this mean-field model to get insights into the microscopic model is called **mean-field analysis (MFA)**.

The most interesting feature of this process is that the mean-field model is not developed as a model for the underlying real-system, but as a model for the microscopic model itself – a model for a model. As a matter of transitivity, developing a (valid) model for a different (valid) model of a real system also results in a (valid) model for the real system. As the microscopic model usually involves much more details than the mean-field model cutbacks in form of simplifications and approximations are necessary. This level of cutback finally decides about the success of the MFA:

- If the level of cutback was too large, the resulting mean-field model is too much simplified. Hence, (some) features of the mean-field model are not features of the microscopic model anymore.
- If the level of cutback was too small, the resulting mean-field model might still be too complex for formal investigation. Thus, a mean-field model is found, but cannot be used.

The so-called **mean-field theory** deals with the question of how to find a suitable mean-field model for a given formalised microscopic model. It is a comparably young science, basically arisen from the field of statistical mechanics, but its area of application has long grown to many other scientific research fields like game-theory, economy, epidemiology and several other social sciences. A couple of researchers that have strongly influenced this science in its early days should be mentioned:

The first ones actually using mean-field models as a supplement for microscopic models were Pierre Curie (1895) and Pierre Weiss (1907), both French physicists, who investigated the average/aggregated spin of ferromagnetic particles. The underlying model, nowadays known

under the term Ising-model formalised by Ernst Ising in 1924 [Ising, 1925], is now one of the most rigorously researched models of statistical physics. Moreover, it is probably the most prominent example for an application of mean-field theory as well. Based on the results of Curie and Weiss many prominent physicists like Einstein, Ehrenfest and Van der Waals continued working on this model and tried to develop answers for its phase-transition problem. Finally Lew Dawidowitsch Landau was able to state a more-less generalised theory for it in 1937 that summarised the ideas of his prominent colleagues [Landau, 1937]. His results can be denoted as the first mean-field theory. For more information about the Ising problem and the related early history of mean-field theory the reader is referred to [Kadanoff, 2013].

Since 1937 a lot of other physicists contributed to that theory. Also the work of N.G. Van Kampen should be mentioned whose Diffusion Approximation for Markov processes published in 1982 [Kampen, N. G. van, 1982, Kampen, 2007] posed motivation for several researchers to apply mean-field methods in other fields of research too. Modern mean-field theory is a very broad scientific field with a lot of different applications. In its heart a couple of generic mathematical theorems can be found which can mostly be attributed to the work of Andrej N. Kolmogorov. This Russian mathematician was one of the most important probabilist of the 20^{th} century and is responsible for most of the probabilistic foundation of modern mean-field theory. His research about Markov processes published in [Kolmogorov, A, 1931] and later in his famous book "Grundbegriffe der Wahrscheinlichkeitsrechnung" [Kolmogorov, 1933] is the basis of most of the mathematical statements about mean-field models.

In this work we aim to show and apply mathematical formulas, so-called **mean-field theorems (MFTs)**, that do not only constitute for conditions under which a macroscopic model poses for a valid mean-field model of a microscopic one, but can also be used to derive one out of the other. We want to underline that these formulas work for formalised, conceptional models and do (basically) not consider the simulation of the model – i.e. we derive a conceptional macroscopic model based on a microscopic ones without execution of any implemented model. We will discuss later, how these methods can be applied vice versa as well.

Unfortunately mean-field theorems require the microscopic models to be mathematically formalised. As mentioned in Chapter 2 this can be quite problematic as they are usually given in a flow-chart or algorithmic notation. Hence, we first state a section that presents formalisation ideas for microscopic models. After that we present a couple of MFTs which can be used to find adequate mean-field models for given microscopic models in case their formalised representation fulfils specific properties. We present two very general theorems and several applied corollaries that can be deduced from them. To prevent confusion with the terminology the reader is invited to take a look at table 4.1

4.1 Formalised Microscopic Model

As mentioned in Chapter 2 problems with theoretic analysis of microscopic models are often caused by a missing formal basic definition. Consequently, we decided to establish a common formal base which is suitable to depict almost all features of microscopic models. One can imagine this base as a formal model for the microscopic model – a "model-model" which will henceforth be denoted as **formalised microscopic model**. In order to obtain this very likely

Term	Meaning
mean-field (of a model)	Term used for the aggregated numbers of a model (compare with Definition 2.1)
mean-field theory	A research field that deals with the question, how (stochastic) microscopic models behave on the aggregated (mean-field) level
mean-field analysis	The analysis process how a microscopic model behaves on the aggregate level
inverse mean-field analysis	A term that will be used by us to describe the analysis process of a macroscopic model using a correlated microscopic approach
mean-field theorem	A theorem that states formal connections between a microscopic and a macroscopic model
mean-field equation	The equation, usually a differential equation, that lies in the heart of a mean-field theorem and poses for the "un-parametrised hull" of the resulting mean-field model.

Table 4.1: Glossary of some mean-field related terminology.

abstraction processes need to be done, as, already mentioned, microscopic models are usually defined in a more algorithmical language based on transitions and flows than a formal language e.g. based on probabilities. Cutback and approximations may be necessary as well, but are not intended in the first place.

In this thesis the most important feature of mentioned formal basis is the idea that the formalised model is beneficial for mean-field analysis as is compatible for being used in mean-field theorems. Unfortunately, most mean-field theorems have quite demanding technical requirements for the mechanisms of the formalised microscopic models. Thus, it might be necessary to perform mentioned cutbacks and approximations to establish a second formalised microscopic model that (a) fulfils the requirements of the mean-field theorem, and (b) is a good approximation to the initial formalised model and the original microscopic model.

Thus, in order to obtain a good macroscopic approximation for a given microscopic model, the modeller needs to perform the following steps:

1. Formalise the given model in the sense of an formalised microscopic model.

2. Probably preform cutbacks and approximations and establish a simplified formalised model that fulfils the prerequisites of a suitable mean-field theorem.

3. Apply the mean-field theorem to receive a mean-field model.

The last part is described in Section 4.6 in details. For the first two steps no general concept can be given, but we will give several examples how this process might work in Chapter 5. The

role of this interposed, additional abstraction process for the mean-field analysis is sketched in Figure 4.1.

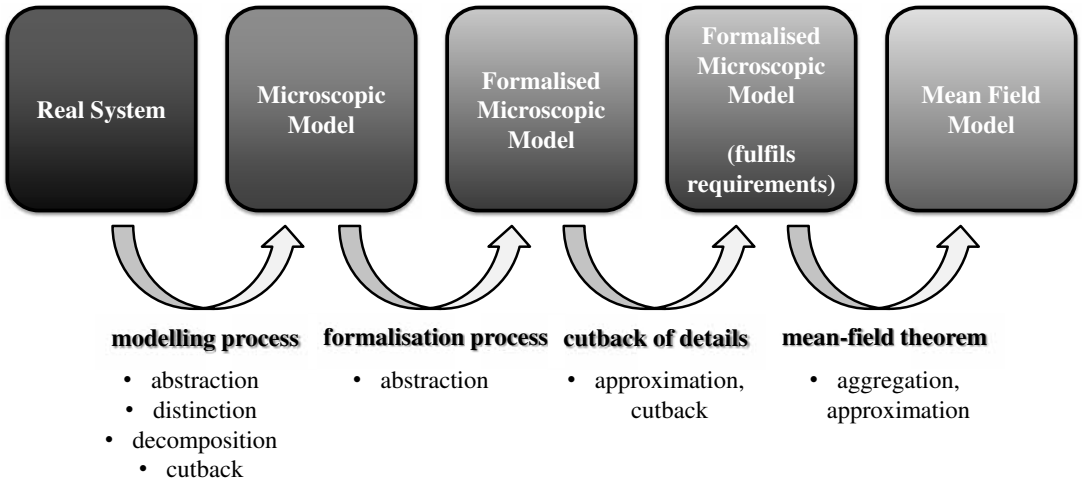


Figure 4.1: Role of the formalised microscopic model for mean-field analysis.

4.1.1 Formalised Microscopic Model for Static Population Models

As classic mean-field theorems in literature are based on population static microscopic models (see Definition 0.3), i.e. models in which the number of interacting sub-models is fixed from the start, we will first state a definition for a formalised microscopic model for static population models. Moreover, the distinction between static and dynamic in this matter makes sense as this definition is a lot more intuitive than the corresponding definition for a dynamic population model given in the next section.

Definition 1.1: Formalised Microscopic Model (Static Population)

First, let $T \subset \mathbb{R}^+$ stand for a continuous or discrete compact time-set. Moreover, define $N \in \mathbb{N}, N \gg 1$ regular, separable stochastic processes (see Definitions 4.8 and 4.4)

$$I_i : (\Omega, T_i) \to \Gamma : (\omega, t) \mapsto I_i(\omega, t) =: I_i(t)(\omega),\ i \in \{1, \dots N\}. \tag{4.1}$$

The processes, henceforth called **agents**, share the same probability, time- and state- space. Moreover, the random process

$$M(t) := \bigtimes_{i=1}^{N} I_i(t) \tag{4.2}$$

is called **model state vector** and is a vector-valued stochastic process with dimension N. Each agent's dynamics are uniquely defined by the transition probabilities

$$P(I_i(t) \in A | M(t_1) = m_1, M(t_2) = m_2, \dots), \tag{4.3}$$

for all sets A, model state vectors $m_1, m_2, \dots$ and all points in time $t > t_1 > \dots > t_m$. Finally the transition distributions may not depend on the agent's index which is guaranteed by

$$P(I_i(t) \in A | M(s_1) = m_1, \dots) = P(I_j(t) \in A | M(s_1) = m_1^{ij}, \dots),\ \forall i, j \in \{1, \dots, N\} \tag{4.4}$$

wherein m_1^{ij} denotes m_1 with switched coordinates i and j. Interchangeability needs to be guaranteed for the initial condition as well

$$P(I_i(t) \in A) = P(I_j(t) \in A), \forall i, j. \tag{4.5}$$

The summary of all these parts is called an **formalised microscopic model** with static population.

On first inspection the definition above does not seem to omit the idea of heterogeneity among the sub-models of a microscopic model as the last condition implies that all agents need to be stochastically interchangeable. To solve this problem, heterogeneity must not be mapped as some static property of the stochastic process but needs to be depicted as an additional dimension to the state-space Γ. Hence, static properties of the individual sub-models need to be abstracted as static states of the corresponding stochastic processes which the process can neither leave nor enter (almost surely). Imagine an agent-based predator-prey model wherein the sub-models are either predators or prey. The correctly modelled formalised microscopic model consists of stochastically equivalent random processes on a state-space that includes the dimension

$$\{0, 1\} \cong \{\text{predator}, \text{prey}\}.$$

The requirement that the transition probabilities of the agents may depend on an arbitrary number of previous states of the total model and not only the state of the agent alone basically

leads to a more general concept than classic stochastic processes. One could describe them as single coordinates of the model state vector via

$$P(I_i(t) \in A | M(t_1) = m_1, M(t_2) = m_2, \dots) \\ = P(M_i(t) \in A | M(t_1) = m_1, M(t_2) = m_2, \dots).$$

This point of view is yet unintuitive. Hence, we prefer the idea of defining the stochastic process of each agent on an extended state-space. Nevertheless this augmentation is (usually) universal enough to cover any kind of defined communication between sub-models in a microscopic model.

Finally neither regularity nor separability of the stochastic process pose for a large constraint regarding generality. Both properties are required not only for the original microscopic model in order to be executable with a computer, but are also observed in any real system: There is no process in reality that changes its state unsteadily an infinite number of times during a finite time interval, and there is no process that needs to remember every single point in its history in order to make a decision how to proceed.

4.1.2 Formalised Microscopic Model for Dynamic Population Models

Clearly, Definition 1.1 is limited when it comes to population-dynamic microscopic models, i.e. models in which the number of sub-models is not fixed from the beginning but may vary during the simulation. In order to cover that we state a definition for an formalised microscopic model for dynamic population. Also more advanced mean-field theorems need to be applied for models defined this way.

Definition 1.2: Formalised Microscopic Model (Dynamic Population)

First, let $T \subset \mathbb{R}^+$ stand for a continuous or discrete compact time-set. Moreover, define regular, separable stochastic processes (see definitions 4.8 and 4.4) via

$$I_i : (\Omega, T_i) \to \Gamma : (\omega, t) \mapsto I_i(\omega, t) =: I_i(t)(\omega),\ i \in \mathbb{N}. \tag{4.6}$$

The processes, called agents, share the same probability and state-space but may differ in time space. Each agent's time space T_i is itself defined by two real valued random variables τ_i^{start} and τ_i^{stop} on $\Omega \times T$ via

$$T_i(\omega) := \{t \in T : \tau_i^{\text{start}}(\omega) \leq t < \tau_i^{\text{stop}}(\omega)\}$$

which we denote as starting and stopping time of I_i as they are representing its life span. Moreover, we define the **model state vector**

$$M(t) := \bigtimes_{i:t \in T_i} I_i(t) \tag{4.7}$$

which is a vector-valued stochastic process with dynamic vector dimension. Each agent's dynamics are uniquely defined by the transition probabilities

$$P(I_i(t) \in A | M(t_1) = m_1, M(t_2) = m_2, \dots), \tag{4.8}$$

for all sets A, model state vectors $m_1, m_2, \dots$ and all points in time $t > t_1 > \dots > t_m$. We conventionally write

$$N(t) := |\{i : t \geq \tau_i^{\text{start}}\}| \tag{4.9}$$

as the total number of agents at a given point in time. The transition distributions may not depend on the agent's index which is guaranteed by

$$P(I_i(t) \in A | M(t_1) = m_1, \dots) = P(I_j(t) \in A | M(t_1) = m_1^{ij}, \dots), \forall i, j \in \{1, \dots, N(t)\} \tag{4.10}$$

wherein m_1^{ij} denotes m_1 with switched coordinates i and j. Interchangeability needs to be guaranteed for the initial condition as well

$$P(I_i(t) \in A) = P(I_j(t) \in A), \forall i, j. \tag{4.11}$$

Finally, the two sets

$$\{t \in T : \tau_i^{\text{start}} \leq t\}, \quad \{t \in T : \tau_i^{\text{stop}} \leq t\} \tag{4.12}$$

need to be measurable with respect to the (natural filtration of the) model state vector $M(t)$.

The additional requirements that $\tau_i^{\text{start}} < \tau_i^{\text{stop}}$ almost surely and

$$P(\tau_i^{\text{start}} \leq t) = 0$$

for almost all i are useful. This requirement might seem confusing but is necessary to avoid infinite numbers of agents. Say $P(\tau_i^{\text{start}} \leq t) > 0$ for an infinite number of indices also an infinite number of starting times will fulfil $\tau_i^{\text{start}}(t) \leq t$ (independent on how small the probability is chosen). Hence, an infinite number of processes starts in a finite time-span which is highly unwanted.

Clearly, the population-dynamic formulation is the more general one and covers the definition of the population-static formalised microscopic model by setting $P(\tau_i^{\text{start}} \leq t) = 1$ and $P(\tau_i^{\text{stop}} < t) = 0 \; \forall i \leq N$ and $P(\tau_i^{\text{start}} \leq t) = 0 \; \forall i > N$.

It is not surprising that definition of the model regards generation of new agents using the concept of a starting time is very inconvenient here because the generation of new sub-models is a delicate problem in population dynamic microscopic models in general. While the removal of sub-models is usually triggered by the sub-model itself (it dies, it gets eaten, it is not needed anymore) the generation of a sub-model cannot be triggered this way as the sub-model does not yet exist. Hence, it can either be triggered by a different sub-model (e.g. via a birth process) or by some superordinate process. We will consider both in the mean-field theorems for dynamic population models.

Finally, the concept of finding a common base for microscopic models via presented formalised microscopic models not only poses the base for the application of mean-field theorems (and comparison on the aggregate level), they moreover make it possible to compar different microscopic models on the microscopic level as well: Stochastic processes can be analysed with fundamental stochastic methods like conditional probabilities/expectations limit theorems or other techniques presented in Chapter 3. We may investigate the influence of specific parameters, analyse difference between time-discrete and time-continuous models, investigate the impact of different time-step sizes etc..

4.2 Introduction to Mean-Field Theorems

First, we discuss, which variables are the objectives of MFA.

4.2.1 Aggregated Numbers

As a direct application of the formalised microscopic model we introduce a formal definition for an important output observable of microscopic models which was already discussed mentioned in Chapter 2.

Definition 2.1: Aggregated Numbers / Mean-Field

Given a formalised microscopic model according to Defintion 1.1 or 1.2 with N (or $N(t)$, respectively) agents I_i at a given point in time t, then

$$X : A_\sigma(\Gamma) \times T \to \mathbb{N} : (A,t) \mapsto X(A,t) = \sum_{i=1}^{N(\text{or } N(t))} \mathbb{1}_A(I_i(t)) \tag{4.13}$$

is denoted as **aggregation mapping** of the model. For fixed A the resulting value is called (time dependent) **aggregated number** of the model.

The aggregated number, also-called **mean-field**, basically counts all agents that share a specific state. Clearly, $X(\Gamma,t) = N$ or $N(t)$ as all agents necessarily have a state in the state-space.

Remark 2.1:

For population-static models it is convenient to investigate the normed aggregated numbers

$$X_0(A,t) := \frac{1}{N} X(A,t) \tag{4.14}$$

as they vary on $[0,1]$ instead of $\{0,\dots,N\}$. This feature simplifies analysis of limits $N \to \infty$ which is a highly interesting feature for the development of mean-field theorems, see in the following chapter. Yet this definition cannot be extended to dynamic population models. Hence, we do not use it as it lacks of generality.

With the stated definition of the aggregation mapping via indicator functions it is clearly interesting whether it is possible to find a **aggregation density** mapping $\tilde{X} : \Gamma \times T \to \mathbb{R}$ so that

$$X(A,t) = \int_A \tilde{X}(x,t)d\mu(x), \quad \forall A \in A_\sigma(\Gamma). \tag{4.15}$$

This feature, although highly wanted, cannot be guaranteed to maintain the universality of the measure space and relates to the Theorem of Radon-Nikodym. For example considering a one-dimensional real-valued state-spaces I_i we would clearly get

$$\tilde{X}(x,t) = \sum_{i=1}^{N(\text{or } N(t))} \delta(x - I_i(t)) \tag{4.16}$$

with the standard delta distribution. As the latter is, being precise, not a function, neither is $\tilde{X}$. We could only imagine it as a generalised function on $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.

Finally, it is important to state that growing N will always lead to growing aggregated numbers. Hence

$$X(A,t) = \mathcal{O}(N),\ \tilde{X}(x,t) = \mathcal{O}(N) \tag{4.17}$$

or

$$X(A,t) = \mathcal{O}(N(0)),\ \tilde{X}(x,t) = \mathcal{O}(N)(0), \tag{4.18}$$

respectively for population dynamic models, as the doubled number of sub-models will lead to doubled aggregated numbers and doubled aggregation density.

As all agents $I_i : \Omega \times T \to \Gamma$ are random numbers, the aggregation mapping, the aggregation density and the aggregated number for a specific set are trajectories of real valued random processes on a common probability space Ω. Hence, it is possible to investigate expected values and variances. This is the main focus of the following topics.

4.2.2 Concepts of Mean-Field Theorems

As stated in [Bicher et al., 2017a], the term mean-field theorem denotes any theorem dealing with the macroscopic approximation of microscopic models. The basic concept of these theorems is shown in Figure 4.2: The deterministic moments of the aggregated numbers (i.e. the mean-field) of a microscopic model are depicted by a so-called mean-field model. This equation-based model, i.e. the shape of the equations and their parametrisation, is directly derived from the formalised microscopic model according to the statement of the mean-field theorem. If applied correctly, a mean-field theorem ensures that the (usually numerically gained) results of the mean-field model are proven to approximate (usually according to some asymptotic equivalence result) the aggregated and statistically processed results of the microscopic model.

Most of these theorems deal with the prediction of the expected value and state asymptotic equivalence to ordinary differential equations (ODEs) or systems of ODEs, respectively (see [Aoki, 2002, de Aguiar et al., 2003, Fibich and Gibori, 2010, Tainaka, 1989, Webb et al., 2007, Nirei, 2006]), but there are also versions using partial differential equations [Deffuant et al., 2000] or even difference equations [Gast and Gaujal, 2011] as well. It depends on the structure of the state-space Γ, the time-space T and whether the model is population-dynamic or -static

which of the theorems can be applied or is most applicable with respect to approximation errors, respectively.

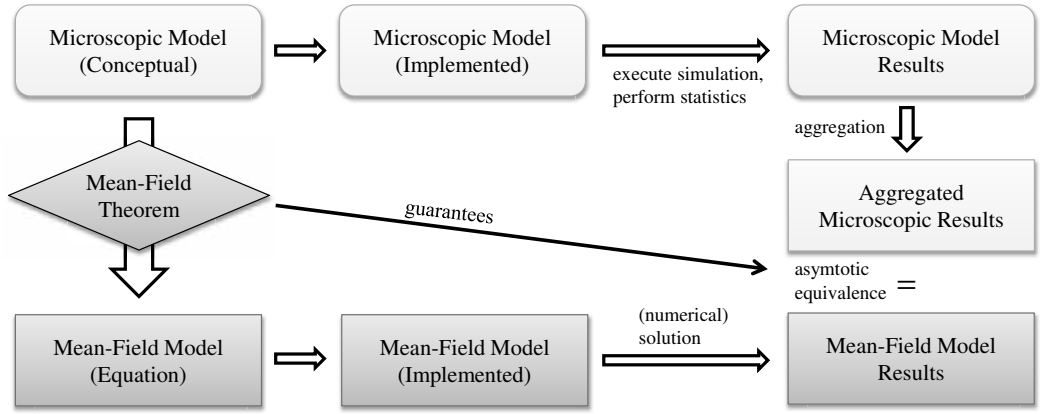


Figure 4.2: General concept of the application of mean-field theorems. This image is originated in [Bicher et al., 2017a]

4.3 General Mean-Field Theorem for Static Population Models

We will start with the most universal mean-field theorem. We hereby refer to the standard mean-field theorem for static population which can also be found in standard literature (e.g. [Tokuzo and Hiroshi, 1985]). Although the statement of the theorem itself is no new contribution to the research field, a new version of the proof was developed by the author in [Bicher, 2016].

As the use of the theorem is quite restrictive we first give a section containing prerequisites the formalised microscopic model has to fulfil to apply the theorem. If the model does not fulfil these properties, either the theorem cannot be used at all, or a more simplified version of the formalised microscopic model has to be established.

4.3.1 Prerequisites for MFTs for Static Population Models

Mean-Field Theorem 3.1 (and consequently all other mean-field theorems in this work that use a static population size, i.e. Corollaries 5.3,5.1 and 5.5) can be applied if a given formalised microscopic model according to Definition 1.2 has the following properties (we use the notation from Definition 1.2 in this section).

- The time set T is a real interval $[t_0, t_{\text{end}}]$. Conventionally we assume $t_0 = 0$.
- All agent processes are memoryless processes according to Definition 4.5 (it is already a precondition of the formalised microscopic model that they are separable and regular).

- Moreover, the probability of the state transitions of an agent may not depend on arbitrary state vectors of the model but **only** on, first, the agent's own state and, second, the aggregated density $X = \tilde{X}$ of the model

$$P(I_i(t) \in A | M(t_1) = m_1) = P(I_i(t) \in A | I_i(t_1) = i_1, X(\cdot, t)). \tag{4.19}$$

  This is probably the greatest restriction to a model and implies that the transition of an agent may not depend on specific states of other individual agents but only on a summary of all of them. Even though it was already mentioned that the aggregated density does not always exists as a real-valued function, it can yet be used as a formal concept to define transition rates.

- The transition-rates

$$\begin{aligned} &\omega : T \times \Gamma \times A_\sigma(\Gamma) \times \Psi \to \mathbb{R} : \\ &t, x, A, X \mapsto \omega(t, s, A, X) = \lim_{h \to 0} \frac{P(I_i(t+h) \in A | I_i(t) = s, X(\cdot, t)) - \mathbb{1}_A(s)}{h} \end{aligned} \tag{4.20}$$

  uniquely describe the process. Herein Ψ denotes the path-space of the aggregation density (we don't want to go into details regarding this space). In summary, agents need to describe regular time-continuous processes.

  We often decide to drop the notation for the dependency on the aggregated number for space reasons

$$\omega(t, s, A) := \omega(t, s, A, X) := \omega(t, s, A, X(\cdot, t))$$

  whenever the type of the used argument is clear by context. Note, that the transition kernel mapping does not depend on the used agent's index as all agents are interchangeable by definition of the formalised microscopic model.

- There exists a transition-kernel $\tilde{\omega}$ in form of a finite function (Compare (4.10)) so that

$$\int_A \tilde{\omega}(t, s_1, s, X) d\mu(s) = \omega(t, s_1, A, X) \tag{4.21}$$

  for all $A \in A_\sigma(\Gamma)$. As mentioned, we write short

$$\omega(t, s_1, s) := \tilde{\omega}(t, s_1, s) := \tilde{\omega}(t, s_1, s, X)$$

  whenever the type of the used argument is clear by context.

- The kernel mapping is differentiable w.r. to its fourth input parameter with bounded first derivative

$$\forall t, s_1, s, X : \quad \left| \frac{\partial}{\partial X} \omega(t, s_1, s, X) \right| < \frac{C}{N} \tag{4.22}$$

  in the sense of the underlying space.

Remark 3.1:

Note, that the derivative needs to vanish linearly when $N \to \infty$ as

$$\begin{aligned}\mathcal{O}(1) &= \mathcal{O}(\omega(t, s_1, s, X) - \omega(t, s_1, s, Y)) \\ &= \mathcal{O}\left(\frac{\partial}{\partial X}\omega(t, s_1, s, \Xi) \cdot (X - Y)\right) = \mathcal{O}\left(\frac{\partial}{\partial X}\omega(t, s_1, s, \Xi)\right)\mathcal{O}(N)\end{aligned}$$

needs to be fulfilled for some sampling point Ξ according to the secant-equation.

- The initial states $I_i(0)$ of the agent processes are a sequence of independent, identically distributed (iid) random numbers on Γ. Their common initial distribution has a density function $f : \Gamma \to \mathbb{R}^+$.

  Thus, for all $A \in A_\sigma(\Gamma)$

$$\begin{aligned}\mathbb{E}\left(X(A, 0)\right) = \mathbb{E}\left(\sum_{i=1}^{N} \mathbb{1}_A(I_i(0))\right) &= \sum_{i=1}^{N} \mathbb{E}\left(\mathbb{1}_A(I_i(0))\right) \\ &= \sum_{i=1}^{N} P(I_i(0) \in A) = \sum_{i=1}^{N} \int_A f(s) d\mu(s) = \int_A N f(s) d\mu(s)\end{aligned}$$

  is fulfilled. Hence, the initial expected value has a density representation.

These requirements seem very demanding at first and exclude broad classes of (population static) formalised microscopic models for applications of mean-field theorems. Yet, we will show that it is, first, possible to relax some of these prerequisites and, second, that it is possible to make useful cutbacks with the given formalised microscopic model without cutting significant model parts.

4.3.2 Statement of the MFT for Static Population Models

Theorem 3.1: General MFT for Static Population

Let

$$\omega = \omega(t, s_1, s, X(\cdot, t))$$

be the transition-kernel of an formalised microscopic model that fulfils all prerequisites from Section 4.3.1 and let $f(s)$ denote the density function of the common initial distribution of $I_i(0)$.

The solution $\varphi(s,t) : \Gamma \times T \to \mathbb{R}^+$ of the differential equation

$$\frac{d}{dt}\varphi(s,t) = \int_\Gamma \varphi(s_1,t)\omega(t,s_1,s,\varphi(\cdot,t)) - \varphi(s,t)\omega(t,s,s_1,\varphi(\cdot,t))d\mu(s_1) \tag{4.23}$$

with

$$\varphi(s,0) = Nf(s) \tag{4.24}$$

fulfils

$$\forall A \in A_\sigma(\Gamma): \quad \mathbb{E}(X(A,t)) = \int_A \varphi(s,t)d\mu(s) + \mathcal{O}(\sqrt{N}).$$

Moreover, for all $A \in A_\sigma(\Gamma)$ and $t \in T$

$$|X(A,t) - \mathbb{E}(X(A,t))| = \mathcal{O}(\sqrt{N}). \tag{4.25}$$

Theorem 3.1 implies that, under some circumstances, the expected value of the aggregated numbers of a formalised microscopic model follows a trajectory that is described by a system of differential equations (coupled by an integral term) with an error that depends on $\sqrt{N}$. As already mentioned that $X(A,t) = \mathcal{O}(N)$ also

$$\mathbb{E}(X(A,t)) = \mathcal{O}(N). \tag{4.26}$$

Thus, the error vanishes asymptotically compared to the scale of X and φ. The second statement can be interpreted as a Law of Large Numbers for interacting stochastic processes. Hence, the solution φ is not only an estimator for $\mathbb{E}(X)$ but also for X which makes it valuable for the analysis of models.

The proof of this theorem is surprisingly troublesome and can, as mentioned, be found in [Bicher, 2016]. It is freely available online. The stated Theorem above matches Theorem 0.4.2 of mentioned document in combination with Theorem 0.6.1 and the first part of Corollary 0.6.1. A very similar Theorem has already been proved by Tokuzo in [Tokuzo and Hiroshi, 1985] but the proof in [Bicher, 2016] uses more fundamental mathematics and does not apply the central-limit theorem of Dynkin and Mandelbaum [Dynkin and Mandelbaum, 1983].

The tricky part of the proof is clearly the interaction condition

$$\omega(t, s_1, s) = \omega(t, s_1, s, X)$$

which makes $\omega(t, I_i(t), s, X)$ a random number not only depending on $I_i(t)$, but on the aggregated number X as well. This problem is overcome using a Taylor expansion (in this case sometimes called Kramers-Moyal expansion [Moyal, 1949]) w.r. to the normed arguments of ω via

$$\omega(t, I_i(t), s, X) \approx \omega(t, I_i(t), s, \mathbb{E}(X)) + \frac{X - \mathbb{E}(X)}{N} \frac{\partial}{\partial X} \omega(t, I_i(t), s, \mathbb{E}(X))$$

As, moreover, the central limit condition $|\mathbb{E}(X(A,t)) - X(A,t)| = \mathcal{O}\left(\sqrt{N}\right)$ is shown to be fulfilled for agents with mean-field interactions, the error terms related to $\frac{X-\mathbb{E}(X)}{N}$ vanish asymptotically with speed $\sqrt{N}$. The central limit expression is proven by showing

$$\mathbb{V}(X) = \mathcal{O}(N)$$

which directly leads to

$$P\left(|\mathbb{E}(X(A,t)) - X(A,t)| \geq N^{\alpha}\right) \leq \frac{\mathbb{V}(X(A,t))}{N^{2\alpha}} \underset{N\to\infty}{\longrightarrow} 0$$

for any $\alpha > \frac{1}{2}$ using the standard Chebyshev inequality.

As mentioned in [Bicher, 2016], it is possible to state a mean-field theorem in form of an integral equation that describes the temporal development for the covariance $\mathrm{Cov}(X(A,t), X(B,t))$ as well. Yet it is hardly applicable for any not solely discrete state-space Γ. Thus, we do not state the general theorem for it in this section, but solely for a discrete state-space Γ in 4.5.1.

4.4 General MFT for Dynamic Population Models

Based on Theorem 3.1 we will establish an analogous mean-field theorem for a formalised dynamic population model according to Definition 1.2. As we could not find any similar approaches in literature the theorem presented in this section can be considered as new.

4.4.1 Prerequisites for MFTs for Dynamic Population Models

To apply the following mean-field theorem for dynamic population models (and all other more applied corollaries for dynamic population models) the prerequisites for the static population model (4.3.1) need to be fulfilled. Clearly, substitution of N by $N(t)$ is required. In addition the starting and stopping times τ_i^{start} and τ_i^{stop} need to be defined by specific rates.

- There is a finite $N(0) = N_0$ for which $\forall\{1, \ldots, N_0\} : \tau_i^{\text{start}} = 0$. All other $\tau_i^{start}, i \geq N_0 > 0$ almost surely.
- The stopping times τ_i^{stop} of all processes need to be uniquely defined by one rate function

  $$d : T \times \Gamma \to \mathbb{R}^+ : (t, s) \mapsto d(t, s) \tag{4.27}$$

  so that for every living agent i

  $$\lim_{h\to 0} \frac{P(\tau_i^{\text{stop}} = t + h | I_i(t) = s)}{h} = d(t, s). \tag{4.28}$$

  Thus, it is the rate that an agent with a state s dies and we will refer to it as **death-rate**.

- The starting times τ_i^{start} are defined by two rate functions. Hereby agents as well as the underlying system may create other agents.

- First there needs to be a rate function

$$c : T \times \Gamma \times A_\sigma(\Gamma) \to \mathbb{R}^+ : (t, s, A) \mapsto c(t, s, A) \tag{4.29}$$

with

$$\forall j : t \geq \tau_j^{\text{start}} : \lim_{h \to 0} \frac{P(\tau_{N(t)+1}^{\text{start}} = t + h, I_{N(t)+1}(t + h) \in A | I_j(t) = s)}{h} = c(t, s, A). \tag{4.30}$$

It defines the rate that an agent with state s creates a new agent with state inside A, hence we will refer to it as **local birth-rate**.

- Analogously there needs to be a function

$$C : T \times A_\sigma(\Gamma) \to \mathbb{R}^+ : (t, A) \mapsto C(t, A) \tag{4.31}$$

with

$$\lim_{h \to 0} \frac{P(\tau_{N(t)+1}^{\text{start}} = t + h, I_{N(t)+1}(t + h) \in A)}{h} N(0) = C(t, A), \tag{4.32}$$

that describes the rate that the overall system creates an agent with state inside A. Multiplication with $N(0)$ ensures that this global creation rate is normed to $N(0)$ and scales correctly, when the number of agents is increased.

We will refer to this rate as **global birth rate**

- Analogously to the transition-rates of the agents, also the two rate functions c, C have kernel representations with, exemplary,

$$\int_A \tilde{c}(t, s_1, s) d\mu(s) = c(t, s_1, A). \tag{4.33}$$

Like in the static-population theorem we drop the $\tilde{\cdot}$ notation whenever it is clear by context that the kernel is used. Note, that d is already defined as a in kernel.

- Moreover, also the three rate/kernel functions d, c, C and hence the correlated probabilities for the starting and stopping times may additionally depend on the current state of the aggregated density $X := \tilde{X}$. Exemplary:

$$d(t, s_1, s) = d(t, s_1, s, X(\cdot, t)). \tag{4.34}$$

As in (4.22), they need to be steadily differentiable w.r. to their fourth parameter with bounded first derivative.

4.4.2 Statement of the MFT for Dynamic Population Models

Theorem 4.1: General MFT for Dynamic Population

Let ω, d, c, C be the transition-, death- and local and global birth-kernels of a formalised microscopic model that fulfils all prerequisites from Section 4.4.1. Moreover, let $f(s)$ denote the density function of the common initial distribution of $I_i(0),\ i \in \{1, \ldots, N(0)\}$.

The solution $\varphi(s,t) : \Gamma \times T \to \mathbb{R}^+$ of the differential equation

$$\frac{d}{dt}\varphi(s,t) = \int_\Gamma \varphi(s_1,t)\omega(t,s_1,s,\varphi(\cdot,t)) - \varphi(s,t)\omega(t,s,s_1,\varphi(\cdot,t))d\mu(s_1) + \int_\Gamma \varphi(s_1,t)c(t,s_1,s,\varphi(\cdot,t))d\mu(s_1) - \varphi(s,t)d(t,s,\varphi(\cdot,t)) + C(t,s) \quad (4.35)$$

with

$$\varphi(s,0) = N(0)f(s) \quad (4.36)$$

fulfils

$$\forall A \in A_\sigma(\Gamma): \quad \mathbb{E}(X(A,t)) = \int_A \varphi(s,t)d\mu(s) + \mathcal{O}(\sqrt{N(0)}).$$

Moreover, for all $A \in A_\sigma(\Gamma)$ and $t \in T$

$$|X(A,t) - \mathbb{E}(X(A,t))| = \mathcal{O}(\sqrt{N(0)}). \quad (4.37)$$

The proof for this theorem is a novel contribution and found in the Appendix A. Analogously to Theorem 3.1 it results with an ODE system that describes the temporal behaviour of the mean-value of the aggregated numbers of a population-dynamic model. Although it seems as if the error bounds for this equation are equivalent to those of the population static equation experiments showed that the error constants C that fulfil $|\varphi(A,t) - \mathbb{E}(X(A,t))| \leq CN(0)$ have to be chosen much higher than for the static population case. The reason lies in the proof of the theorem which is based on the comparison between a dynamic-population model and a corresponding static-population model with an additional state *idle* and a larger (static) total population of $N' := N(t) + N_i(t)$. Hereby N' does not only include all $N(t)$ agents in the original dynamic model but also all $N_i(t)$ *idle* agents. If only a small amount of agents are initially not *idle* while a lot of agents are generated, i.e. switch from *idle* to a not-*idle* state, during the simulation then $N_i(0)$ is large and $N' \gg N(0)$. Hence, $|\varphi(A,t) - \mathbb{E}(X(A,t))| \leq CN' = C'N(0)$ with $C' \gg C$.

It is a reasonable question why a microscopic model with dynamic population is not abstracted (for mean-field analysis purposes) by a population static model with an idle state in the first place. First, it is unnatural to keep microscopic sub-processes in the model that are in an absorbing state like death or some state that is not regarded within the model boundaries. Second, it is not possible to give a good estimate for initially required agents to cover all creation or birth processes that will be required for the entire simulation – at least not with analytic means.

In the following section we aim to rephrase the Mean-Field Theorems 5.1 and 5.2 for a number of specific state-spaces Γ. A couple of corollaries result which are much more applicable than the two general theorems.

4.5 Mean-Field Equations for Specific State Spaces

Given the two key theorems of this thesis we state some special cases for them that are useful for specific applications.

4.5.1 Mean-Field Equations for Discrete State Space

For the first special case we consider $\Gamma := \{s_1, s_2, \dots, s_d\}$. Clearly, all formulas in this section also hold for infinite state-spaces like $\Gamma = \mathbb{N}$ or $\Gamma = \mathbb{Z}$ when applying the limit $d \to \infty$, as long as this is possible w.r. to finiteness of the sums (i.e. integrability with respect to the counting measure).

As the state-space is discrete it is possible to write the aggregated numbers $X(A,t)$ as sums of vector elements $\vec{X}(t)$ with

$$\vec{X}_j(t) := \sum_{i=1}^{N \text{ or } N(t)} \mathbb{1}_{\{s_j\}}(I_i(t)), \quad X(A,t) = \sum_{s_i \in A} \vec{X}_i(t). \tag{4.38}$$

This vector matches the aggregated density of the model as the integral with respect to the counting measure is equivalent to a sum. We refer to this as the **aggregated vector** or mean-field vector of the model. Moreover, the transition kernel can be written as a matrix

$$\omega(t,i,j,\vec{X}) := \omega(t,s_i,\{s_j\},\vec{X}(t)) \tag{4.39}$$

which may depend on elements of the aggregated vector. Note, that death- and birth- kernels d, c, C for dynamic population models can be defined as matrices and vectors, as well.

Corollary 5.1: Discrete Space MFT for Static Population

Let $\omega(t,i,j,\vec{X}(t)) := \omega(t,s_i,s_j,\vec{X}(t))$ be the transition-kernel of a formalised microscopic model that fulfils all prerequisites from Section 4.3.1 with $\Gamma := \{s_1,\dots,s_n\}$ and let $\vec{f} \in \mathbb{R}^n$ denote the probabilities of the common initial distribution of any $I_j(0)$ via $\vec{f}_i = P(I_j(0) = s_i)$.

The solution vector $\vec{\varphi}(t) : T \to (\mathbb{R}^+)^n$ of the differential equation system

$$\frac{d}{dt}\vec{\varphi}_i(t) = \sum_{j=1}^{n} \vec{\varphi}_j(t)\omega(t,j,i,\vec{\varphi}(t)) - \vec{\varphi}_i(t)\omega(t,i,j,\vec{\varphi}(t)) \tag{4.40}$$

with

$$\vec{\varphi} = N\vec{f} \tag{4.41}$$

fulfils

$$\mathbb{E}(\vec{X}(t)) = \vec{\varphi}(t) + \mathcal{O}(\sqrt{N}).$$

Above version of the theorem is probably the most prominent version of a mean-field theorem as most microscopic models are designed on discrete state-spaces. In the easiest case of $\Gamma = \{0, 1\}$ the proof is very simple and can be found in [Aoki, 2002]. As a direct extension we state the analogous theorem for the dynamic-population case.

Corollary 5.2: Discrete Space MFT for Dynamic Population

Let ω, d, c, C be the transition-, death- and local and global birth- kernels of a formalised microscopic model that fulfils all prerequisites from Section 4.4.1 with $\Gamma := \{s_1, \ldots, s_n\}$. Moreover, let $\vec{f} \in \mathbb{R}^n$ denote the probabilities of the common initial distribution of any $I_j(0)$ via $\vec{f}_i = P(I_i(0) = s_i)$.

The solution vector $\vec{\varphi}(t) : T \to (\mathbb{R}^+)^n$ of the differential equation system

$$\begin{aligned}\frac{d}{dt}\vec{\varphi}_i(t) = &\sum_{j=1}^{n} \vec{\varphi}_j(t)\omega(t, j, i, \vec{\varphi}(t)) - \vec{\varphi}_i(t)\omega(t, i, j, \vec{\varphi}(t)) \\ &+ \sum_{j=1}^{n} \vec{\varphi}_j(t)c(t, j, i, \vec{\varphi}(t)) - \vec{\varphi}_i(t)d(t, i, \vec{\varphi}(t)) + C(t, i, \vec{\varphi}(t)) \quad (4.42)\end{aligned}$$

with

$$\vec{\varphi}(0) = N(0)\vec{f} \quad (4.43)$$

fulfils

$$\mathbb{E}(\vec{X}(t)) = \vec{\varphi}(t) + \mathcal{O}(\sqrt{N(0)}).$$

Proof. Both theorems are proved when applying Theorem 3.1 and Theorem 4.1 for a discrete state-space with n elements and the counting measure $\mu(A) = |A|$. ■

4.5.2 Mean-Field Equations for Continuous State Space

In this section we aim to apply the Mean-Field Theorems 3.1 and 4.1 to gain mean-field equations for continuous sets $\Gamma \subseteq \mathbb{R}^n$ with the Lebuesgue measure λ. We will write $\vec{q}$ instead of $\vec{s_1}$ in the mean-field theorems to avoid that $\vec{s_1}$ is confused with the first coordinate of $\vec{s}$ and not a second state besides $\vec{s}$.

As it would be redundant to state equations (4.23) and (4.35) simply replacing $d\mu(s) = d\lambda(\vec{s})$ we decided to go one step further:

It is a valid assumption in microscopic models with continuous states, that the solution of the mean-field equation can at least be approximated by a density function $\varphi(\vec{s}, t),\ s \in \mathbb{R}^n$ which is differentiable with respect to space. Thus, we may use Taylor expansion of $\varphi(\vec{q}, t)$ w.r. to space at $\varphi(\vec{s}, t)$ in

$$\int_\Gamma \varphi(\vec{q}, t)\omega(t, \vec{q}, \vec{s}) - \varphi(\vec{s}, t)\omega(t, \vec{s}, \vec{q})d\mu(\vec{q}).$$

It is legit to cut this Taylor expansion after one or two elements in case $\omega(t, \vec{s}, \vec{q})$ vanishes with large differences of $|\vec{s} - \vec{q}|$. As a result, the terms $\varphi(\vec{s}, t)$ and its derivatives can be dragged out of the integral terms and a partial differential equation (PDE) results.

Corollary 5.3: Continuous Space MFT for Static Population

Let $\omega(t, \vec{s}, \vec{s}', X(\cdot, t))$ be the transition-kernel of a formalised microscopic model that fulfils all prerequisites from Section 4.3.1 with an open, convex $\Gamma \subset \mathbb{R}^n$. We additionally require that $\omega(t, \vec{s}, \vec{s}', X(\cdot, t))$ vanishes for $|\vec{s} - \vec{s}'| > L$ for some small positive L. Moreover, let $f(\vec{s})$ denote the density of the initial distribution of any $I_j(0)$ via $\int_A f(\vec{s})d\vec{s} = P(I_j(0) \in A)$.

The (probably weak) solution $\varphi(\vec{s}, t) : \Gamma \times T \to \mathbb{R}$ of the partial differential equation

$$\frac{\partial \varphi}{\partial t}(\vec{s}, t) = \varphi(\vec{s}, t)F1(t, \vec{s}, \varphi) + \sum_{i=1}^{n} \frac{\partial \varphi}{\partial \vec{s}_i}(\vec{s}, t)F2_i(t, \vec{s}, \varphi) + \\ + \sum_{i,j=1}^{n} \frac{\partial^2 \varphi}{\partial \vec{s}_i \vec{s}_j}(\vec{s}, t)F3_{i,j}(t, \vec{s}, \varphi), \quad (4.44)$$

with

$$F1(t, \vec{s}, \varphi) := \int_{\mathbb{R}^n} \omega(t, \vec{q}, \vec{s}, \varphi(\cdot, t)) - \omega(t, \vec{s}, \vec{q}, \varphi(\cdot, t))d\vec{q},$$
$$F2_i(t, \vec{s}, \varphi) := \int_{\mathbb{R}^n} \omega(t, \vec{q}, \vec{s}, \varphi(\cdot, t))(\vec{q}_i - \vec{s}_i)d\vec{q},$$
$$F3_{i,j}(t, \vec{s}, \varphi) := \frac{1}{2}\int_{\mathbb{R}^n} \omega(t, \vec{q}, \vec{s}, \varphi(\cdot, t))(\vec{q}_i - \vec{s}_i)(\vec{q}_j - \vec{s}_j)d\vec{q},$$

and

$$\varphi(0, \vec{s}) = Nf(\vec{s}) \quad (4.45)$$

fulfils

$$\forall A \in A_\sigma(\mathbb{R}^n) : \quad \mathbb{E}(X(A, t)) = \int_A \varphi(\vec{s}, t)d\vec{s} + \mathcal{O}(\sqrt{N}) + E(\omega) \quad (4.46)$$

wherein $E(\omega)$ is an error term that depends on the condition of the rate function ω w.r. to the size of possible jumps.

Proof. As mentioned we use the Taylor expansion (which requires convexity of the area for the remainder)

$$\varphi(\vec{s}\,',t) = \varphi(\vec{s},t) + \sum_{i=1}^{d}(\vec{s}\,'-\vec{s})_i \frac{\partial\varphi(\vec{s},t)}{\partial\vec{s}_i} + \frac{1}{2}\sum_{i,j=1}^{n}(\vec{s}\,'-\vec{s})_i(\vec{s}\,'-\vec{s})_j \frac{\partial^2\varphi(\vec{s},t)}{\partial\vec{s}_i\vec{s}_j} + \mathcal{O}\left(|\vec{s}\,'-\vec{s}|^3\right).$$

While inserting the first Taylor parts into the general mean-field equation (4.23) leads to the stated formula (4.44), the remainder leads to the integral

$$\int_{\mathbb{R}^n} \mathcal{O}\left(|\vec{q}-\vec{s}|^3\right)\omega(t,\vec{q},\vec{s},\varphi(\cdot,t))d\vec{q} =: E(\omega) \tag{4.47}$$

which we assume to be small, as ω is defined to have limited support w.r. to $|\vec{q}-\vec{s}|$. ■

Analogously we receive the following corollary for the population-dynamic case.

Corollary 5.4: Continuous Space MFT for Dynamic Population

Let ω, d, c, C be the transition-, death- and local and global birth- kernels of a formalised microscopic model that fulfils all prerequisites from Section 4.4.1 with an open, convex $\Gamma \subset \mathbb{R}^n$. We additionally require that $\omega(t, \vec{s}, \vec{s}', X)$ and $c(t, \vec{s}, \vec{s}', X)$ vanish for $|\vec{s} - \vec{s}'| > L$ for some small positive L. Moreover, let $f(\vec{s})$ denote the density of the initial distribution of any $I_j(0)$ via $\int_A f(\vec{s})d\vec{s} = P(I_j(0) \in A)$.

The (probably weak) solution $\varphi(\vec{s}, t) : \Gamma \times T \to \mathbb{R}$ of the partial differential equation

$$\frac{\partial \varphi}{\partial t}(\vec{s}, t) = C(t, \vec{s}) + \varphi(\vec{s}, t)F1(t, \vec{s}, \varphi) + \sum_{i=1}^{n} \frac{\partial \varphi}{\partial \vec{s}_i}(\vec{s}, t)F2_i(t, \vec{s}, \varphi) + \sum_{i,j=1}^{n} \frac{\partial^2 \varphi}{\partial \vec{s}_i \vec{s}_j}(\vec{s}, t)F3_{i,j}(t, \vec{s}, \varphi), \quad (4.48)$$

with

$$F1(t, \vec{s}, \varphi) := \int_{\mathbb{R}^n} \omega(t, \vec{q}, \vec{s}, \vec{u}) + c(t, \vec{q}, \vec{s}, \varphi) - \omega(t, \vec{s}, \vec{q}, \varphi)d\vec{q} - \vec{d}(t, \vec{s}, \varphi),$$
$$F2_i(t, \vec{s}, \varphi) := \int_{\mathbb{R}^n} (\omega(t, \vec{q}, \vec{s}, \varphi) + c(t, \vec{q}, \vec{s}, \varphi)) (\vec{q}_i - \vec{s}_i)d\vec{q},$$
$$F3_{i,j}(t, \vec{s}, \varphi) := \frac{1}{2} \int_{\mathbb{R}^n} (\omega(t, \vec{q}, \vec{s}, \varphi) + c(t, \vec{q}, \vec{s}, \varphi)) (\vec{q}_i - \vec{s}_i)(\vec{q}_j - \vec{s}_j)d\vec{q},$$

and

$$\varphi(0, \vec{s}) = N(0)f(\vec{s}) \quad (4.49)$$

fulfils

$$\forall A \in A_\sigma(\mathbb{R}^n) : \quad \mathbb{E}(X(A, t)) = \int_A \varphi(\vec{s}, t)d\vec{s} + \mathcal{O}(\sqrt{N}) + E(\omega, c) \quad (4.50)$$

wherein $E(\omega, c)$ is an error term that depends on the condition of the rate functions ω and c w.r. to the size of possible jumps.

Proof. The corollary is proven analogously to 5.3. Surprisingly only c, but not d and C, contributes to the error due to the Taylor expansion. ■

It is not surprising that the aggregated numbers of microscopic continuous models can be described with a Fokker-Planck-like equation, yet it is important to note that these equations might be non-linear ones as $F1, F2$ and $F3$ (may) depend on φ as well. It is not clear in the first place if these model equations have a strong or weak solution at all, but we could not find an application for which these equations could not be solved at least numerically. Deeper analysis of these equations would definitely be necessary to underline this statement, but this is not focus of this work.

4.5.3 Mean-Field Theorems for Hybrid State Space

The last version of the classic Mean-Field Theorem we rephrase Theorem 3.1 for population static microscopic models with agents on a hybrid state-space. Hereby we denote the product space of a continuous and a discrete space. That means, in specific,

$$\Gamma := U \times V, \quad U \subseteq \mathbb{R}^n, V := \{s'_1, s'_2, \ldots, s'_{n'}\} \tag{4.51}$$

for two positive $d, d' \in \mathbb{N}$ and U open and convex. Consequently, we adjust Γ with the product measure of the standard Lebesgue measure λ^n and the counting measure. With this measure integrals calculate as follows. With $A \in U$ and $B \in V$ we get

$$\int_{A \times B} f(x) d\mu(x) = \sum_{s'_i \in B} \int_A f(\vec{s}, s'_i) d\vec{s}, \tag{4.52}$$

wherein $x = (\vec{s}, s')$ denotes a vector of the product space $U \times V$. Moreover, any function $f(x) = f((\vec{s}, s'))$ on Γ can be expressed as a vector $\vec{f}(\vec{s})$ of functions on U with $\vec{f}_i(\vec{s}) = f((\vec{s}, s'_i))$. We make use of this observation applied on the aggregated numbers/density and state a mean-field theorem.

Corollary 5.5: Hybrid Space MFT for Static Population

Let

$$\omega(t,(\vec{s},s_i'),(\vec{q},s_j'),\vec{X}),\quad i,j\in\{1,\dots,n'\},\ \vec{s},\vec{q}\in U\subseteq\mathbb{R}^n$$

be the transition-kernel of an formalised microscopic model that fulfils all prerequisites from Section 4.3.1 with additional requirement that $\omega(t,(\vec{s},s_i'),(\vec{q},s_j'),\vec{X}(\cdot,t))$ vanishes for $|\vec{s}-\vec{q}|>L$ for some small positive L. Moreover, let $\vec{f}(\vec{s})$ denote a vector valued function describing the initial distribution any $I_j(0)$ via

$$\int_A \vec{f}_i(\vec{s})d\vec{s}=P(I_j(0)\in A\times\{s_i'\})\quad ,A\subseteq U,s_i'\subseteq V.$$

The (possibly weak) solution $\vec{\varphi}(\vec{s},t):U\times T\to\mathbb{R}^{n'}$ of the partial differential equation system

$$\frac{\partial\vec{\varphi}_i}{\partial t}(\vec{s},t)=\sum_{j=1}^{n'}\vec{\varphi}_j(\vec{s},t)F1^{ji}(\vec{s},\vec{\varphi},t)-\vec{\varphi}_i(\vec{s},t)F1^{ji}(\vec{s},\vec{\varphi},t)+$$
$$\sum_{j=1}^{n'}\left(\sum_{k=1}^{n}\frac{\partial\vec{\varphi}_j}{\partial\vec{s}_k}(\vec{s},t)G_k^{ji}(\vec{s},\vec{\varphi},t)+\sum_{k,l=1}^{n}\frac{\partial^2\vec{\varphi}_j}{\partial\vec{s}_k\vec{s}_l}(\vec{s},t)H_{kl}^{ji}(\vec{s},\vec{\varphi},t)\right),\quad(4.53)$$

with

$$F1^{ji}(\vec{s},\vec{\varphi},t):=\int_U\omega(t,(\vec{q},s_j'),(\vec{s},s_i'),\vec{\varphi})d\vec{q}$$
$$F2^{ji}(\vec{s},\vec{\varphi},t):=\int_U\omega(t,(\vec{s},s_i'),(\vec{q},s_j'),\vec{\varphi})d\vec{q}$$
$$G_l^{ji}(\vec{s},\vec{\varphi},t):=\int_U\omega(t,(\vec{q},s_j'),(\vec{s},s_i'),\vec{\varphi})(\vec{q}_l-\vec{s}_l)d\vec{q}$$
$$H_{lk}^{ji}(\vec{s},\vec{\varphi},t):=\frac{1}{2}\int_U\omega(t,(\vec{q},s_j'),(\vec{s},s_i'),\vec{\varphi})(\vec{q}_l-\vec{s}_l)(\vec{q}_k-\vec{s}_k)d\vec{q},$$

and

$$\vec{\varphi}_i(0,\vec{s})=N\vec{f}_i(\vec{s})\quad(4.54)$$

fulfils

$$\forall A\in A_\sigma(\mathbb{R}^n),\forall i\in\{0,\dots,n'\}:\ \mathbb{E}(X((A,\{s_i'\}),t))=\int_A\vec{\varphi}_i(\vec{s},t)d\vec{s}+\mathcal{O}(\sqrt{N})+E(\omega)\quad(4.55)$$

wherein $E(\omega)$ is an error term that depends on the condition of the rate function ω w.r. to the size of possible jumps.

Proof. First writing the integral in the general Mean-Field Theorem 3.1 as sum of integrals

according to (4.52) leads the corresponding integral formulation of (4.53). Afterwards, applying a Taylor approximation analogously to the proof of Corollary 5.3 leads the stated PDE system. ■

Surprisingly the hybrid formulation of the Mean-Field Theorem is more useful than the solely continuous one. Reason for that might be that there is hardly any microscopic model that has no discrete state. Note, that any kind of heterogeneity among population has to be mapped onto the state-space according to the definition of the formalised microscopic model. Simple examples which cause those kind of discrete states based on heterogeneity are male/female, predator/prey or healthy/infected.

Corollary 5.6: Hybrid Space MFT for Dynamic Population

Let ω, d, c, C be the transition-, death- and local and global birth- kernels of a formalised microscopic model that fulfils all prerequisites from Section 4.4.1 with additional requirement that

$$\omega(t, (\vec{s}, s_i'), (\vec{q}, s_j'), \vec{X})$$

and

$$c(t, (\vec{s}, s_i'), (\vec{q}, s_j'), \vec{X})$$

vanish for $|\vec{s} - \vec{q}| > L$ for some small positive L. Moreover, let $\vec{f}(\vec{s})$ denote a vector valued function describing the initial distribution of any $I_j(0)$ via

$$\int_A \vec{f}_i(\vec{s}) d\vec{s} = P(I_j(0) \in A \times \{s_i'\}) \quad , A \subseteq U, s_i' \subseteq V.$$

We will not state the resulting PDE system, as already the population-static version of the hybrid space MFTs resulted in unhandy, long equations. It is a direct mixture of the population-dynamic mean-field equation for continuous space (4.48) and the population-static mean-field equation for hybrid space (4.53). One may imagine the resulting PDE system as system (4.48) wherein each line contains an additional sum with respect to the discrete part of the state-space.

4.5.4 Variance Theorem for Discrete State Space

As final mean-field theorem of this section we will state and proof a theorem that can be used to supplement the Discrete Space MFT for Static Population 5.1. In contrast to the mean-field theorems presented before this theorem cannot be applied to find a mean-field model for the expected value, but it can be used to find a measure, how well the mean-field model given by the Discrete Space MFT for Static Population represents the mean-field trajectories of the microscopic model.

Theorem 5.1: Discrete Space Variance Theorem for Static Population

Assuming that the prerequisites and the notation of the Discrete Space MFT for Static Population 5.1 hold with the additional requirement that $\omega(t,i,j,\vec{X})$ has finite second derivatives w.r. to $\vec{X}$, then the solution vector $(\xi_{i,j}(t))_{i,j\in\{1,\dots,n\}}$ of

$$\begin{aligned}\frac{d}{dt}\xi_{i,j}(t) = &-\vec{\varphi}_i(t)\omega(t,i,j,\vec{\varphi}(t)) - \vec{\varphi}_j(t)\omega(t,j,i,\vec{\varphi}(t))\\ &+ \mathbb{1}_i(j)\sum_{k=1}^{n}\vec{\varphi}_k(t)\omega(t,k,i,\vec{\varphi}(t)) + \vec{\varphi}_i(t)\omega(t,i,k,\vec{\varphi}(t))\\ &+ \sum_{k=1}^{n}\xi_{i,k}(t)\omega(t,k,j,\vec{\varphi}(t)) - \xi_{i,j}(t)\omega(t,j,k,\vec{\varphi}(t))\\ &+ \sum_{k=1}^{n}\xi_{j,k}(t)\omega(t,k,i,\vec{\varphi}(t)) - \xi_{j,i}(t)\omega(t,i,k,\vec{\varphi}(t))\\ &+ \sum_{m=1}^{n}\xi_{i,m}(t)\sum_{k=1}^{n}\left(\vec{\varphi}_k(t)\frac{\partial\omega(t,k,j,\vec{\varphi}(t))}{\partial\vec{X}_m} - \vec{\varphi}_j(t)\frac{\partial\omega(t,j,k,\vec{\varphi}(t))}{\partial\vec{X}_m}\right)\\ &+ \sum_{m=1}^{n}\xi_{j,m}(t)\sum_{k=1}^{n}\left(\vec{\varphi}_k(t)\frac{\partial\omega(t,k,i,\vec{\varphi}(t))}{\partial\vec{X}_m} - \vec{\varphi}_i(t)\frac{\partial\omega(t,i,k,\vec{\varphi}(t))}{\partial\vec{X}_m}\right)\end{aligned} \tag{4.56}$$

wherein $\vec{\varphi}(t)$ stands for the solution vector of (4.40) and

$$\xi_{i,j}(0) = \mathrm{Cov}(\vec{X}_i(0),\vec{X}_j(0)) = N\left(\mathbb{1}_i(j)\vec{f}_i - \vec{f}_i\vec{f}_j\right) \tag{4.57}$$

fulfils

$$|\xi_{i,j}(t) - \mathrm{Cov}(\vec{X}_i(t),\vec{X}_j(t))| = \mathcal{O}(\sqrt{N}). \tag{4.58}$$

The proof for this theorem can be found in the appendix via A.0.2 where it is directly proven for the given state-space and is based on Corollary 5.1. In [Bicher, 2016] a more simplified Theorem can be found for a general state-space. Yet, the proof there is long, complicated and is only valid for small first derivatives $\nabla\omega$.

The variance theorem does not seem to be practicable from the point of application, yet it is very interesting when analysing the resulting system of differential equation. We state, without showing the proof here, a few facts about the ODE system. All of them can be verified with simple, yet time-consuming, calculations.

- The initial values clearly fulfil

$$\xi_{i,i}(0) = \mathbb{V}(\vec{X}_i(0)) > 0, \quad i\neq j:\ \xi_{i,j}(0) = \mathrm{Cov}(\vec{X}_i(0),\vec{X}_j(0)) < 0. \tag{4.59}$$

  This can be reasoned as the sum of all agents needs to be equal to N. Hence, the higher one of the aggregated numbers, the more likely are the others to be lower – they are negatively

correlated. Surprisingly there is no condition in the equation (for $n > 2$) that this remains fulfilled for any $t > 0$. Hence, it might be possible that one of the aggregated numbers in the model actually correlates positively with a second one although the sum of all needs to be N all the time.

- The solution of the variance equation fulfils a lot of properties that hold for covariances of random numbers in general. We list a few:

  1. **Symmetry:** $\mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t)) = \mathrm{Cov}(\vec{X}_j(t), \vec{X}_i(t))$ holds as
  $$\xi_{i,j}(0) = \xi_{j,i}(0) \quad \text{and} \quad \frac{d\xi_{i,j}}{dt}(t) = \frac{d\xi_{j,i}}{dt}(t). \tag{4.60}$$
  2. **Summing to zero I:** $0 = \mathbb{V}(N) = \mathbb{V}\left(\sum_{i=1}^n \vec{X}_i(t)\right) = \sum_{i,j=1}^n \mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t))$ is fulfilled as
  $$\sum_{i,j=1}^n \xi_{i,j}(0) = 0 \quad \text{and} \quad \sum_{i,j=1}^n \frac{d\xi_{i,j}}{dt}(t) = 0. \tag{4.61}$$
  3. **Summing to zero II:** More generally for any two index-sets I, J with $I \cup J = \{1, \dots, n\}$ the identity
  $$\begin{aligned}\sum_{i,j\in I} \mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t)) &= \mathbb{V}\left(\sum_{i\in I} \vec{X}_i(t)\right) \\ &= \mathbb{V}\left(N - \sum_{i\in J} \vec{X}_i(t)\right) = \mathbb{V}\left(\sum_{i\in J} \vec{X}_i(t)\right) = \sum_{i,j\in J} \mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t))\end{aligned}$$
  is fulfilled by
  $$\sum_{i,j\in I}^n \xi_{i,j}(0) = \sum_{i,j\in J}^n \xi_{i,j}(0) \quad \text{and} \quad \sum_{i,j\in I}^n \frac{d\xi_{i,j}}{dt}(t) = \sum_{i,j\in J}^n \frac{d\xi_{i,j}}{dt}(t). \tag{4.62}$$

- These three formulas can now help to reduce the order of the system as not all equations are really necessary. For $\Gamma = \{s_1, \dots, s_n\}$ the original system would suggest that n^2 coupled differential equations need to be solved to find a solution of the system. Note, that we could imagine the solutions of the system as a matrix
$$\begin{pmatrix} \xi_{1,1} & \xi_{1,2} & \dots & \xi_{1,n} \\ \xi_{2,1} & \xi_{2,2} & \dots & \xi_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{n,1} & \xi_{n,2} & \dots & \xi_{n,n} \end{pmatrix}.$$
First, the symmetry makes it possible to reduce this matrix to upper (or lower) triangular form as $\xi_{i,j} = \xi_{j,i}$
$$\begin{pmatrix} \xi_{1,1} & \xi_{1,2} & \dots & \xi_{1,n} \\ 0 & \xi_{2,2} & \dots & \xi_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \xi_{n,n} \end{pmatrix}.$$

This way, the system is fully described by $\frac{n(n+1)}{2}$ equations. Clearly, the second identity can be seen as a special case of the third one with I being the empty set, hence we will not regard it any longer. Regarding the third identity it seems as if a large number of constraints make it possible to reduce the system heavily – there are 2^{n-1} constraints defined by the third identity

$$\sum_{i,j\in I} \xi_{i,j} = \sum_{i,j\in J} \xi_{i,j}.$$

Unfortunately, it turns out that all equations of this shape linearly depend on the n equations

$$\sum_{i,j\in\{1,\dots,k-1,k+1,\dots,n\}} \xi_{i,j} = \xi_{k,k}.$$

Thus, only n of possible 2^{n-1} equations can be reduced without loss of information.

It finally results that the n^2 equations of (4.56) can be reduced to $\frac{n(n+1)}{2} - d = \frac{n(n-1)}{2}$ equations.

■ While the covariances $\xi_{i,j}(t), i \neq j$ are, in general, of minor concern for the analysis of a microscopic model, the variances $\xi_{i,i}(t)$ are much more valuable. We could use them to establish confidence intervals for the spread of the simulation results. Hence, it might be interesting to know if it is possible to derive an ODE system that only solves for the variance curves and does not require any covariances.

Based on previous calculations the system may be reduced to $n(n-1)/2$ equations which suits to the sequence $n(n-1)/2 = 1, 3, 6, 10, \dots$ for $n = 2, 3, 4, 5 \dots$. As a state-space with n elements leads to n-dimensional aggregated number vector and n variance curves, only for $n = 2$ and $n = 3$ a system of equations that only consists of variances but not of covariances can be found, as $n \geq n(n-1)/2$. We state the resulting systems for $n = 2$ and $n = 3$:

For $\Gamma := \{s_1, s_2\}$ we get

$$\begin{aligned}\frac{d}{dt}\xi_{1,1}(t) &= \vec{\varphi}_2(t)\omega(t,2,1,\vec{\varphi}) + \vec{\varphi}_1(t)\omega(t,1,2,\vec{\varphi}) - 2\xi_{1,1}(t)(\omega(t,1,2,\vec{\varphi}) + \omega(t,2,1,\vec{\varphi})) \\ &+ \xi_{1,1}(t)\left(\vec{\varphi}_1\left(\partial_2\omega(t,1,2,\vec{\varphi}) - \partial_1\omega(t,1,2,\vec{\varphi})\right) + \vec{\varphi}_2\left(\partial_1\omega(t,2,1,\vec{\varphi}) - \partial_2\omega(t,2,1,\vec{\varphi})\right)\right)\end{aligned} \tag{4.63}$$

Only one equation is needed as $n(n-1)/2 = 1$. Note, that $\xi_{1,1}(t) = \xi_{2,2}(t)$ based on the third stated identity.

For $\Gamma := \{s_1, s_2, s_3\}$ we get

$$\forall i \neq j \neq k : \ \xi_{i,j} = \frac{\xi_{k,k} - \xi_{i,i} - \xi_{j,j}}{2}$$

and therefore the three differential equations are

$$\begin{aligned}\frac{d}{dt}\xi_{j,j}(t) = &\sum_{i=1,i\neq j}^{3}\left(\vec{\varphi}_i(t)\omega(t,i,j,\vec{\phi}) + \vec{\varphi}_j(t)\omega(t,j,i,\vec{\phi})\right)\\ &+\sum_{i=1,i\neq j,k\notin\{i,j\}}^{3}\left(\xi_{k,k}(t)-\xi_{i,i}(t)-\xi_{j,j}(t)\right)\omega(t,i,j,\vec{\phi}) - 2\xi_{j,j}(t)\sum_{i=1,i\neq j}^{3}\omega(t,j,i,\vec{\phi})\\ &+\sum_{i=1,i\neq j,k\notin\{i,j\}}^{3}\left(\xi_{k,k}(t)-\xi_{i,i}(t)-\xi_{j,j}(t)\right)\sum_{l=1}^{3}\left(\vec{\varphi}_l(t)\partial_i\omega(t,l,j,\vec{\phi}) - \vec{\varphi}_j(t)\partial_i\omega(t,j,l,\vec{\phi})\right)\\ &+2\xi_{j,j}(t)\sum_{k=1}^{3}\left(\vec{\varphi}_k(t)\partial_j\omega(t,k,j,\vec{\phi}) - \vec{\varphi}_j(t)\partial_j\omega(t,j,k,\vec{\phi})\right)\end{aligned} \tag{4.64}$$

with $j \in \{1,2,3\}$.

4.6 Usage of Mean-Field Theorems

4.6.1 General Applicability of Mean-Field Theorems

With exception of the variance theorem, all mean-field theorems presented in the last section are developed to establish a "valid" mean-field model for a given microscopic system. Here "valid" means, that the results of the mean-field model (can be proven to) approximate the results of the microscopic model for a sufficiently large number of individual sub-models. Yet, the methods do not state any kind of interpretation of the newly generated model, as the model structure and the model parameters are solely gained by formal transformations and calculations and, in principle, not by a modelling process.

From this point of view it is worthy to discuss whether the gained ODE/PDE system is a "model" or not. Based on the definitions of Stachowiak [Stachowiak, 1973], a model needs to be a simplified image of a natural or artificial system, which is not given in the first place. Yet we might find reasonable explanations for the validity of the mean-field model as a model for the real system afterwards, by identifying relations and causalities. For this process we cannot give any formulas and methods.

The stated mean-field theorems are, in principle, developed to derive a macroscopic model from a given microscopic model by proving their equivalence on the asymptotic level. Hence, the theorems also state that the results of a microscopic model match the results of a given macroscopic one, if the transition rates are chosen according to the parameters of the mean-field equation. Thus, we can clearly derive microscopic models from given macroscopic ones as well if the macroscopic models can be matched with one of the mean-field equations. We denote this process as **inverse mean-field analysis**.

Summarising, the following purposes for application of mean-field theorems can be detected:

1. **Find a supplemental, probably simplified, macroscopic model that can be used to replace the microscopic one.** Evaluation times or unwanted stochasticity might for example motivate these intentions. Hereby we need to care about relationships between reality and the mean-field model and we need to identify how and why specific parts of the real-system lead to which parts of the mean-field model. Any necessary approximation and abstraction that was required to fulfil the properties of the mean-field theorems need to be done with the real-system in mind as it might result in terrible modelling errors in the macroscopic model otherwise.

2. **Find equations that describe the aggregated behaviour of a given microscopic model for analysis of the microscopic model.** Hereby we do not care about whether or not the derived equations can be seen as a model for the real system and we do not attempt to supplement the microscopic model by the mean-field model. We are solely interested in understanding the qualitative/quantitative behaviour of the microscopic model based on the macroscopic one which might be easier to investigate.

3. **Compare two given models, a microscopic and a macroscopic one.** In case a macroscopic model can be transformed to fulfil a mean-field equation then the corresponding microscopic model can be compared with the given one. Vice versa, if a microscopic model is capable of applying a mean-field theorem, we could compare the resulting mean-field model with a given macroscopic one. The latter is probably easier as a macroscopic simulation usually executes faster and is easier to analyse.

4. **Find a microscopic description of a given macroscopic model for analysis purposes.** Hereby we could think of validation of the macroscopic model: Do the defined processes in the macroscopic description make sense if analysed on the microscopic level? Probably also difficult numerical properties of the macroscopic system, e.g. in case of highly sensitive and stiff ODEs or PDEs, could motivate this idea, when trying to find a microscopic "solver" of the macroscopic equation.

5. **Find a supplemental, probably simplified, microscopic model that can be used to replace the macroscopic one.** This process is usually motivated when given macroscopic models need to be extended by microscopic rules which are unhandy or, in some cases, even impossible to be modelled in a macroscopic way.

Although the purposes in 1/2 and 4/5 seem very similar at first the differences are fundamental. Quick mean-field analysis as done for purposes 2 and 4 can be used for **verification** of either of the models (compared to the other) while only when done with care, as for 1 and 5, the newly derived model can be used for **validation** purposes.

4.6.2 Step By Step from Microscopic to Macroscopic

Especially for finding a macroscopic analogue for microscopic model sometimes tricky abstraction and approximation processes are necessary to fulfil the prerequisites of a mean-field theorem. Hereby it is necessary to carefully decide which parts of the model cannot be neglected as

they are relevant for the system behaviour and which parts can. Note, that it is not always possible to find a suitable mean-field model (using a mean-field theorem) that represents the system behaviour of the microscopic model.

As there is hardly any microscopic simulation model that is defined on a fully formalised basis it turns out that the key to find a good mean-field model lies within **asking the right questions in correct order**. Thus, we established step-by-step instructions, one for the standard mean-field approximation and one for the inverse mean-field approximation, that guide through the process by asking questions about the microscopic/macroscopic model. Each question will lead to the next question and will result in a valid mean-field model if answered and computed properly. At some stages we can only give hints on how to continue as the process can only be generalised up to some extent. In order to inspire the reader examples on how the process is used will be presented in the next chapter.

Question 2.1:

Identify the individual sub-models of your microscopic model and note their total number $N(0)$ after the initialisation phase of the model. We will henceforth denote them as **agents**.

How many agents $N(0)$ can be identified?	
< 50	Very likely the model is not suited for mean-field analysis. Continue with Question 2.2 only with care.
≥ 50	Continue with Question 2.2.

Question 2.2:

Investigate the behaviour of the agents and summarise all possible states - i.e. find a space Γ that contains (an abstraction) of all possible states of all agents. Note, that any heterogeneity and memory of all agents needs to be mapped onto the state-space to make mean-field analysis possible via the proposed theorems. Moreover, make a "list" of all possible transitions from one state to another that are possible. Add to this list all possibilities that agents in specific states are created or die.

To make sure, that the state-space is chosen large enough:

Think of exchaning the states of any two agents I_i, I_j in your model at any given time: Does this have an influence on the model?	
Yes	Heterogeneity among the agents needs to be mapped onto the state-space. Probably a new state-space dimension is needed. Continue with Question 2.2.
No	Continue with Question 2.3

Question 2.3:

To make sure, that the state-space is chosen large enough:

Does any model action that leads to creation/destruction of sub-models or transition between states depend/rely on any agent property, that is not (yet) included in the state-space?	
Yes	In case this property cannot be neglected as it has too much influence on the microscopic model dynamics it is necessary to add this feature as an additional dimension to the state-space. Continue with Question 2.2.
No	Continue with Question 2.4

Question 2.4:

To make sure, that the state-space is chosen large enough:

For each element in the list: is the state-transition or birth-, death-process influenced by any states of the model that lie in the past?	
Yes, the transition is influenced by the previous k states of the model	Any memory has to be mapped onto the state-space. In case a state s has a memory of k states it is split into k-states with memory 1 (compare with the ideas below Definition 4.5). Continue with Question 2.2.
Yes, the transition is influenced an unknown number of previous states	Any memory has to be mapped onto the state-space. If this is not possible (compare with the ideas below Definition 4.5) then the model is probably not suited for our mean-field analysis.
No	Continue with Question 2.5

Question 2.5:

Finally, distinguish between different state-space types.

How does the common state-space Γ look like?	
It is finite.	Identify $\Gamma = \{s_1, \ldots, s_n\}$. Continue with Question 2.6.
It is a product space and has at least one dimension that is part of the real-axis.	Identify sets $U \in \mathbb{R}^n$ and $V = \{s'_1, \ldots, s'_{n'}\}$ so that $\Gamma = U \times V$ and each element $s \in \Gamma = (\vec{s}, s'_i)$. We will refer to the two vector entries as "continuous-" and "discrete- state" of the agent. Continue with Question 2.6.
It is infinite but discrete.	Identify $\Gamma = \lim_{n\to\infty}\{s_1, \ldots, s_n\}$. Either cut this set after some sufficiently large n (probably some states cannot be reached) or let n run to infinity after the mean-field analysis. For the remainder this analysis process, fix $n < \infty$. Continue with Question 2.6.
None of the above	Either the model is not suited for mean-field analysis or some more abstraction is required. Continue with Question 2.2.

Question 2.6:

Investigate the way, how the individual agents change their states and how the model is updated. Note, t_{end} as the designated end-time of the model.

When are the states of the sub-models updated?	
Simultaneously at equidistant discrete time steps.	Note, $T = (0, t_1, \ldots, t_{\text{end}})$ as the vector of time-steps and Δt as the length of the time interval. Continue with Question 2.7
Simultaneously at fixed non-equidistant discrete time steps.	Note, $T = (0, t_1, \ldots, t_{\text{end}})$ as the vector of time-steps and Δt_i as the length of each time interval. Whenever Δt is addressed in the following steps, replace it with the corresponding Δt_i for the regarded time-step accordingly. Continue with Question 2.7.
Individually with events that take place with specific rates.	Note, $T = [0, t_{\text{end}}]$ and Δt as a small positive real number (compared to t_{end}). Henceforth let $\Delta t \to 0$ whenever this is possible to transform a difference quotient to a differential quotient. Continue with Question 2.7.

Question 2.7:

We start establishing the corresponding mean-field equation by step-by-step noting equations from left to right. Herein we implicitly apply the stated mean-field theorems, wherein we relax a couple of requirements which will be justified in Section 4.6.3.

What is the state-space?	
$\{s_1, \ldots, s_n\}$**.**	Start by noting d left hand sides of differential equations: For all $i \in \{1, \ldots, n\}$ write: $$\frac{d\varphi_i}{dt}(t) = \ldots$$ It is sometimes helpful to use symbols instead of $1, \ldots, n$ as it is more intuitive. In this instruction we will address them as "equation number i". Continue with Question 2.8.
$U \times \{s'_1, \ldots, s'_{n'}\}$**.**	Start by noting n' left hand sides of partial differential equations: For all $i \in \{1, \ldots, n'\}$ write: $$\frac{\partial\varphi_i}{\partial t}(\vec{s}, t) = \ldots$$ It is sometimes helpful to use symbols instead of $1, \ldots, n'$ as it is more intuitive. In this instruction we will address them as "equation number i". Continue with Question 2.17.

Question 2.8:

Consider each element of the state-space as an iterative process starting by s_1. Each time this question is reached a fictional counter i should be enhanced. As soon as $i = d + 1$ is reached: Continue with Question 2.16.

Investigate state-space element s_i and look through the list of transitions:

Is it possible that an agent directly switches from state s_i to a different state during the simulation?	
Yes	Search your transition list for any s_j for which a transition $s_i \to s_j$ is possible. Continue with Question 2.9.
No	Continue with Question 2.10.

Question 2.9:

For each j for which the transition $s_i \to s_j$ is possible at some time during the simulation perform the following steps:

For $t \in T$ calculate the corresponding transition probabilities $P(I(t+\Delta t) = s_j | I(t) = s_i)$.

Is it possible to find an approximation so that $$P(I(t+\Delta t) = s_j \mid I(t) = s_i) = f(t, i, j, \vec{X})$$ **wherein** $\vec{X}$ **stands for the vector of aggregated numbers of the model (see 2.1), without changing the system behaviour?**	
Yes	Calculate $$\omega(t, i, j, \vec{X}) := \frac{f(t, i, j, \vec{X})}{\Delta t}$$ as a function of t and $\vec{X}$. Search in the notes for equation number i and add $$-\varphi_i(t)\omega(t, i, j, \vec{\varphi}(t))$$ to the right hand side. Search in the notes for equation number j and add $$+\varphi_i(t)\omega(t, i, j, \vec{\varphi}(t))$$ to the right hand side. In case all j with transitions $s_i \to s_j$ are dealt with: Continue with Question 2.10.
No	Either the state-space was chosen too small or the model is not suitable for application of any of the mean-field theorems presented earlier. Anyway, specific steps have to be reconsidered.

Question 2.10:

Now consider possible creation processes of agents with state s_i.

Is it possible that agents with state s_i **can be created during the simulation?**	
Yes	Continue with Question 2.11.
No	Continue with Question 2.14.

Question 2.11:

Now we need to think about **how** agents with state s_i are created in the microscopic model. Reconsider all states s_j:

Is it possible that an agent with state s_j creates (gives birth) to an agent with state s_i?	
Yes	Search for all states s_j so that an agent in state s_j may create an s_i agent. Continue with Question 2.12.
No	Continue with Question 2.13.

Question 2.12:

For each j for which an agent with state s_j may create one in s_i perform the following steps:

For $t \in T$ calculate the probability $P(\text{New agent in } s_i \text{ after } \Delta t | I(t) = s_j)$.

Is it possible to find an approximation so that

$$P(\textbf{New model in } s_i \textbf{ after } \Delta t | I(t) = s_j) = f(t, j, i, \vec{X}),$$

wherein $\vec{X}$ stands for the vector of aggregated numbers of the model (see 2.1), without changing the system behaviour?

Yes	Calculate $$c(t, j, i, \vec{X}) := \frac{f(t, j, i, \vec{X})}{\Delta t}$$ as a function of t and $\vec{X}$. Search in the notes for equation number i and add $$+\varphi_j(t) c(t, i, j, \vec{\varphi}(t))$$ to the right hand side. In case all processes that an agent creates a new agent are processed, continue below.
No	Either the state-space was chosen too small or the model is not suitable for application of any of the mean-field theorems presented earlier. Anyway, specific steps have to be reconsidered.

Are there any causes for a creation of agents in state s_i which have not been considered yet?	
Yes	Continue with Question 2.13
No	Continue with Question 2.14

Question 2.13:

For $t \in T$ calculate the probability $P(\text{New agent in } s_i \text{ after } \Delta t|t)$ that an agent is created by some other reason than a birth process of a different agent.

Is it possible to find an approximation so that $$P(\textbf{New agent in } s_i \textbf{ after } \Delta t|\textbf{time} = t) = f(i,t,\vec{X})$$ **wherein $\vec{X}$ stands for the vector of aggregated numbers of the model (see 2.1), without changing the system behaviour?**	
Yes	Calculate $$C(i,t,\vec{X}) := \frac{f(i,t,\vec{X})}{\Delta t}$$ as a function of t and $\vec{X}$. Search in the notes for equation number i and add $$+C(i,t,\vec{\varphi}(t))$$ to the right hand side. Continue with Question 2.14.
No	Either the state-space was chosen too small or the model is not suitable for application of any of the mean-field theorems presented earlier. Anyway, specific steps have to be reconsidered.

Question 2.14:

Finally, consider death processes.

Is it possible that an agent with state s_i dies?	
Yes	Continue with Question 2.15.
No	Continue with Question 2.8.

Question 2.15:

For $t \in T$ calculate the probability $P(\text{dead after } \Delta t, I(t) = s_i)$ that an agent dies.

Is it possible to find an approximation so that

$$P(\textbf{dead after } \Delta t, I(t) = s_i) = f(i, t, \vec{X})$$

wherein $\vec{X}$ stands for the vector of aggregated numbers of the model (see 2.1), without changing the system behaviour?

Yes	Calculate $$d(i, t, \vec{X}) := \frac{f(i, t, \vec{X})}{\Delta t}$$ as a function of t and $\vec{X}$. Search in the notes for equation number i and add $$-\varphi_i(t) d(i, t, \vec{\varphi}(t))$$ to the right hand side. Continue with Question 2.8.
No	Either the state-space was chosen too small or the model is not suitable for application of any of the mean-field theorems presented earlier. Anyway, specific steps have to be reconsidered.

Question 2.16:

The mean-field equation should now be finally ready to use. It remains to find a suitable initial condition. For each i with state $s_i \in \Gamma$:

Is the total number $\vec{X}_i(0)$ of agents initially in state s_i deterministic?	
Yes	Set $\varphi_i(0) := \vec{X}_i(0)$
No	Calculate the expected value $\mu_i := \mathbb{E}(\vec{X}_i(0))$ and set $\varphi_i(0) := \mu_i$.

The process is finished. The noted system of differential equations is a valid mean-field approximation for the given microscopic model if all simplifications have been done with care.

Question 2.17:

Now, consider all elements s'_i of the discrete part V of the state-space as an iterative process starting by s'_1. Each time this question is reached, a fictional counter i is enhanced. As soon as $i = n' + 1$ is reached: Continue with Question 2.25.

Look through the list of possible transitions:

Is it possible that an agent directly switches from $(\vec{s}, s'_i)$ to a different state?	
Yes	Search your transition list for all $(\vec{s}_2, s'_j)$ for which a transition $(\vec{s}, s'_i) \to (\vec{s}_2, s'_j)$ is possible. Continue with Question 2.18.
No	Continue with Question 2.19.

Question 2.18:

For each j (including $j = i$) for which the transition $(\vec{s}, s'_i) \to (\vec{q}, s'_j)$ is possible, perform the following steps:

For $t \in T$ calculate a transition density function

$$\tilde{P} : (\vec{q}, t) \mapsto \tilde{P}(I(t + \Delta t) = (\vec{q}, s'_j) | I(t) = (\vec{s}, s'_i))$$

based on the probabilities

$$P(I(t + \Delta t) \in (A, \{s'_j\}) | I(t) = (\vec{s}, s'_i)).$$

This calculation might require the use of the delta distribution.

Is it possible to find an approximation so that

$$\tilde{P}(I(t+\Delta t) = (\vec{q}, s'_j) | I(t) = (\vec{s}, s'_i)) = f((i, \vec{s}), (j, \vec{q}), t, X)$$

wherein X stands for the aggregated density of the model (see 2.1), without changing the system behaviour?

Yes	Calculate $$\omega((i, \vec{s}), (j, \vec{q}), t, X) := \frac{f((i, \vec{s}), (j, \vec{q}), t, X)}{\Delta t}$$ as a function of t and X. Search in the notes for equation number i and add $$-\int_U \varphi_i(\vec{s}, t)\omega((i, \vec{s}), (j, \vec{q}), t, \varphi)d\vec{q}$$ to the right hand side. Search in the notes for equation number j and add $$+\int_U \varphi_i(\vec{q}, t)\omega((i, \vec{q}), (j, \vec{s}), t, \varphi)d\vec{q}$$ to the right hand side. As soon as all possible states with transitions from $(\vec{s}, s'_i)$ are dealt with: Continue with Question 2.19.
No	Either the state-space was chosen too small or the model is not suitable for application of any of the mean-field theorems presented earlier. Anyway, specific steps have to be reconsidered.

Question 2.19:

Now consider possible creation processes of agents with discrete state s'_i.

Is it possible that agents with state $(\vec{s}, s'_i,)$ can be created during the simulation?	
Yes	Continue with Question 2.20.
No	Continue with Question 2.23.

Question 2.20:

Now we need to think about **how** agents with state $(\vec{s}, s'_i)$ are created in the microscopic model.

Is it possible that an agent creates (gives birth) to an agent with state $(\vec{s}, s'_i)$?	
Yes	Make a list of all discrete states s'_j for which an agent with discrete state s'_j may create an $(\vec{s}, s'_i)$ agent. Continue with Question 2.21.
No	Continue with Question 2.22.

Question 2.21:

For each j (including i) for which an agent with discrete state $(\vec{s}', s'_j)$ may create one in $(\vec{s}, s'_i)$ perform the following steps:

For $t \in T$ calculate the transition density function

$$\tilde{P} : (\vec{s}, t) \mapsto \tilde{P}(\text{New agent in } (\vec{s}, s'_i) \text{ after } \Delta t | I(t) = (\vec{q}, s'_j))$$

based on the probabilities

$$P(\text{New agent in } (A, \{s'_j\}) \text{ after } \Delta t | I(t) = (\vec{q}, s'_j)).$$

This calculation might require the use of the delta distribution.

Is it possible to find an approximation so that

$$\tilde{P}(\textbf{New agent in } (\vec{q}, {s'}_j) \textbf{ after } \Delta t | I(t) = (\vec{q}, {s'}_j)) = f((\vec{s}, i), (\vec{q}, j), t, X)$$

wherein X stands for the aggregated density of the model (see 2.1), without changing the system behaviour?

Yes	Calculate $$c((\vec{q}, j), (\vec{s}, i), t, X) := \frac{f((\vec{q}, j), (\vec{s}, i), t, X)}{\Delta t}$$ as a function of t and X. Search in the notes for equation number i and add $$+ \int_U \varphi(\vec{q}, t) c((\vec{q}, j), (\vec{s}, i), t, \varphi) d\vec{q}$$ to the right hand side. As soon as all processes that an agent creates a new agent are dealt with, continue below.
No	Either the state-space was chosen too small or the model is not suitable for application of any of the mean-field theorems presented earlier. Anyway, specific steps have to be reconsidered.

Are there any causes for a creation of agents with discrete state s_i which have not been considered yet?

Yes	Continue with Question 2.22
No	Continue with Question 2.23

Question 2.22:

For $t \in T$ calculate the transition density function

$$(\vec{s}, t) \mapsto \tilde{P}(\text{New agent in } (\vec{s}, {s'}_i) \text{ after } \Delta t)$$

based on the probabilities

$$P(\text{New agent in } (A, \{{s'}_i\}) \text{ after } \Delta t | t)$$

that an agent is created by some other reason than a birth process of a different agent. This calculation might require the use of the delta distribution.

Is it possible to find an approximation so that

$$\tilde{P}(\textbf{New agent in } (\vec{s}, {s'}_i) \textbf{ after } \Delta t) = f((\vec{s}, i), t, X)$$

wherein X stands for the aggregated density of the model (see 2.1), without changing the system behaviour?

Yes	Calculate $$C((\vec{s}, i), t, X) := \frac{f((\vec{s}, i), t, X)}{\Delta t}$$ as a function of t and X. Search in the notes for equation number i and add $$+C((\vec{s}, i), t, \vec{\varphi})$$ to the right hand side. Continue with Question 2.23.
No	Either the state-space was chosen too small or the model is not suitable for application of any of the mean-field theorems presented earlier. Anyway, specific steps have to be reconsidered.

Question 2.23:

Finally, consider death processes.

Is it possible that an agent with state $(\vec{s}, {s'}_i)$ dies?	
Yes	Continue with Question 2.24.
No	Continue with Question 2.8.

Question 2.24:

For $t \in T$ calculate the probability density

$$\tilde{P} : (\vec{s}, t) \mapsto \tilde{P}(\text{dead after } \Delta t, I(t) = (\vec{s}, s'_i))$$

based on the probabilities

$$P(\text{dead after } \Delta t, I(t) \in (A, \{{s'}_i\}))$$

that an agent dies.

Is it possible to find an approximation so that

$$\tilde{P}(\textbf{dead after } \Delta t, I(t) = (\vec{s}, s_i')|\dots) = f((\vec{s}, i), t, X)$$

wherein X stands for the aggregated density of the model (see 2.1), without changing the system behaviour?

Yes	Calculate $$d((\vec{s}, i), t, X) := \frac{f((\vec{s}, i), t, X)}{\Delta t}$$ as a function of t and X. Search in the notes for equation number i and add $$-\varphi(\vec{s}, t)d((\vec{s}, i), t, \varphi)$$ to the right hand side. Continue with Question 2.17.
No	Either the state-space was chosen too small or the model is not suitable for application of any of the mean-field theorems presented earlier. Anyway, specific steps have to be reconsidered.

Question 2.25:

A n'-dimensional mean-field equation was derived in an integral form.

Is the integral representation sufficient?	
Yes	Continue below
No	For all terms $\varphi(\vec{q}, t)$ on the right hand side of the equation, apply a Taylor decomposition at $\varphi(\vec{s}, t)$ via $$\varphi(\vec{q}, t) = \varphi(\vec{s}, t) + (\vec{q} - \vec{s})\nabla\varphi(\vec{s}, t) + \dots$$ This way, these terms can be dragged out of the integrals and a partial-differential system is achieved.

It remains to find a suitable initial condition. Therefore, calculate the initial density

$$\tilde{P}(I(0) = (\vec{s}, s_i'))$$

based on the probabilities

$$P(I(0) \in (A, \{s_i'\})).$$

Set

$$\varphi_i(\vec{s}, 0) := N(0) \cdot \tilde{P}(I(0) = (\vec{s}, s_i')).$$

A mean-field model was derived that, in case all simplifications have been done with care, approximates the mean-field behaviour of the microscopic model.

4.6.3 Scientific Background of the Micro-to-Macro Instruction

Before we state the inverse mean-field analysis instruction, i.e. a step-by-step instruction how macroscopic models can be approximated by microscopic models based on the stated mean-field theorems, we give some **scientific explanation and justification** for the process described in Section 4.6.2. Hereby we focus on describing how the mean-field theorems stated in the sections before are implicitly used and which elements have been relaxed in order to fulfil their requirements.

The basic structure of the process is gained from rough analysis of the shape of the mean-field equations that result from any of the stated mean-field theorems. The easiest way to see this is analysing the structure of the discrete mean-field equation (4.40): Any transition of the microscopic model from one state to an other results in a multiplication of the corresponding mean-value trajectory of the a-priori state multiplied with the corresponding transition kernel. This term is subtracted from the equation of the a-prior state and added to the equation of the a-posterior state. Even though the mean-field theorems for continuous and hybrid state-spaces have a slightly different shape, this basic idea holds. Similar ideas can be applied to death and creation processes too.

We focus on specific relaxations compared to the theorems. The most obvious relaxation between the described mean-field analysis process and the requirements of the mean-field theorems is found in Question 2.6 the difference regarding the simulation time update. While the mean-field theorems require formalised microscopic models which are evaluated with specific transition rates on the continuous time-set $T = [0, t_{\text{end}}]$, the process is defined for discrete-time models as well – in fact, as mentioned, these models are the most common type of microscopic models. Hereby we make use of the idea that a continuous-time microscopic model with specific transition/death/creation rates can be approximated by a discrete-time model with specific transition/death/creation probabilities. Let Y be a time-discrete Markov-process with transition probabilities P and time-steps of length Δt. Define the transition rates ω of a time-continuous process Y' via

$$\forall s \in [t, \Delta t) : \omega_2(s, Y_0, Y_1, \dots) := \frac{P(Y(t + \Delta t) = Y_1 | Y(t) = Y_0, \dots)}{\Delta t}$$

then the time-continuous process Y' will approximate the time-discrete process. As the switch of two states within a time-step is impossible for the time-discrete process, the approximation is good if the probability that the time-continuous process has two state-changes in the time interval Δt is as small. As the probability for two state-changes in a time-continuous Markov chain qualitatively depends on the probability of one state-change squared (compare with the Poisson property 3.63), the approximation is good if the quotient

$$\frac{P(\text{two state-changes})}{P(\text{one state-change})} \approx \frac{P^2}{P} = P$$

has small values, which is guaranteed if the transition probabilities of the time-discrete model are small. This again is influenced by the time-step length, as reducing the latter, also the transition probabilities become smaller. We will discuss this in more detail in Section 5.1 and the reader is referred to this case study for deeper insights.

Consequently, the approximation of a time-discrete microscopic model by a time-continuous model with transformed transition rates via $\omega = P/\Delta t$ is legit, if the transition probabilities for the single sub-models are comparably small which directly correlates with the idea of choosing a small step-size Δt. In case the model uses non-equidistant time-steps, of course, the maximum time-step needs to be investigated. Experiments showed that transition probabilities are very rarely too large for application of the mean-field analysis and that the influence of too large transition probabilities is quite small, compared to too little numbers of sub-models N or $N(0)$, respectively.

In the questions concerning the spatially continuous transition/death and creation rates, we completely neglected the condition that the transition kernel needs to be finite and differentiable functions w.r. to the aggregated density. We even mentioned that we might be required to use the delta distribution to find a transition probability: Clearly, infinite transition kernels pose a direct violation of the prerequisites of any mean-field theorem, but as the real-valued delta distribution can be written as a limit of differentiable finite functions, we may apply the mean-field theorem for any of those *discretisations*. As the delta distributions which might be required in the definition of the rates during the process **never** lead to an irregular process as this would violate in principle a proper definition of a microscopic model, the *discretised* microscopic model would clearly converge towards the given one in probability. Hence, the mean-field approximation is legit for the original model as well.

4.6.4 Step By Step from Macroscopic to Microscopic

In this section we want to give a step-by-step instruction of the inverse mean-field analysis process - i.e. finding a microscopic model, that, observed on the mean-field level, behaves like a given macroscopic model. As for the classic process, we will guide through this process by asking the correct questions in correct order. It will soon turn out that we are able to use much more mathematical arguments than in the classic direction.

The process will result in transition rates and birth/death processes for the microscopic sub-models. Though a microscopic approach in form of a formalised microscopic model is uniquely given, it is clear that, given a set of transition rates ω (birth/death rates) for the individuals, there are various ways how to finally set up a conceptual and finally executable microscopic model. They could e.g. be implemented in a time-discrete agent-based model with time-steps Δt using $P = \omega \cdot \Delta t$ as transition probabilities or they could be used directly in a microscopic discrete event model. Moreover, there are no restrictions on how the derived probabilities/rates result from the dynamics of the microscopic model. They could result from contact processes or from directly implemented algebraic formulas. Hence, there are hardly limits for creative ideas.

We will restrict this process to macroscopic models that consist of, or can be approximated by, systems of ordinary differential equations. These models result in microscopic representations with discrete states. The process is difficult to be generalised for partial differential equations which would lead to spatially continuous models. We will only add a few comments for them afterwards.

Question 4.26:

First, consider the basic shape of the given system.

Is it possible to find a representation

$$\frac{\vec{x}}{dt} = \vec{F}(t, \vec{x}), \quad \vec{x}(0) = \vec{x}_0.$$

of the given model as an explicit system of first order (Note, that any ODE system with higher order can be written as system of first order)?

Yes	Fix $\vec{F}$ as the focus of investigation and note n as the total number of equations (the length of $\vec{F}$). Moreover, define $$\Gamma := \{s_1, \ldots, s_n\}.$$ Continue with Question 4.27
No	The given ODE system is probably not suited for our mean-field analysis as it is implicit.

Question 4.27:

Translation: As a mean-field of a microscopic model is always a positive number it is required to determine lower bound for the system vector.

Can we find a finite real number $0 \geq \alpha > -\infty$ and, if necessary, an end-time $t_{\text{end}} \leq \infty$, so that

$$\forall t \in [0, t_{\text{end}}] : \min(\vec{x}(t)) \geq \alpha?$$

Yes	Define $T = [0, t_{\text{end}}]$ and set $\vec{y} := \vec{x} + (1, \ldots, 1)^T \alpha$. Consequently, the system translates to $$\frac{d}{dt}\vec{y}(t) = \vec{F}(t, \vec{y} - \alpha)$$ with $\vec{y}(0) := \vec{y}_0 := \vec{x}_0 + (1, \ldots, 1)^T \alpha$. Continue with Question 4.28
No	The given ODE system is probably not suited for our mean-field analysis as it is unstable.

Question 4.28:

Scaling: Now, choose a scaling factor $\beta \in \mathbb{R}^+$ so that

$$N(0) = \beta \left\| \vec{y}_0 \right\|_1$$

is a reasonable large number if used as number of initial sub-models. We recommend $N(0) \geq 100$ as $\sqrt{N(0)}$ poses for a measure of fluctuation of the microscopic model. Fix

$$\vec{f} := \frac{1}{\left\| \vec{y}_0 \right\|_1} \vec{y}_0$$

as vector-valued initial distribution of the sub-model states Γ via

$$P(I(0) = s_i) = \vec{f}_i.$$

Moreover, set $\vec{z} := \vec{y}\beta$ and calculate

$$\frac{d}{dt}\vec{z}(t) = \beta \vec{F}\left(t, \frac{\vec{y} - \alpha}{\beta}\right) := \vec{H}(t, \vec{z})$$

with $\vec{z}_0 := \beta(\vec{x}_0 + (1, \ldots, 1)^T \alpha)$. Continue with Question 4.29.

Question 4.29:

Next, write each row of $\vec{H}$ in the form

$$\vec{H}_i(t, \vec{z}) =: \sum_{j=1}^{q_i} h_{ij}(t, \vec{z})$$

and study the summands. First of all, we need to sort them w.r. to their sign. For all $i \in \{1, \ldots n\}$ and $j \in \{1, \ldots, q_i\}$:

Consider $\vec{z}$ as a vector of positive numbers and $t \in [0, t_{\text{end}}]$. What is the sign of $(t, \vec{z}) \mapsto h_{i,j}(t, \vec{z})$?	
Always positive	Leave $h_{i,j}(t, \vec{z})$ unchanged. Continue with Question 4.30.
Always negative	Set $h_{i,j}(t, \vec{z}) := -h_{i,j}(t, \vec{z})$ and change its sign in the sum above to $-$. Continue with Question 4.30.
Depends on $\vec{z}$ or t	Split the summand into a positive and a negative part hereby increasing q_i by one. Set $$h_{i,k+1} := h_{i,k}$$ for all $k > j$, $$h_{i,j}(t, \vec{z}) := \max(h_{i,j}(t, \vec{z}), 0)$$ and $$h_{i,j+1}(t, \vec{z}) := \max(-h_{i,j}(t, \vec{z}), 0)$$ and change the sign of the second summand in the sum above to $-$. Continue with Question 4.30.

Question 4.30:

As a result of the last question each row of the right-hand-side function should have the form

$$\vec{H}_i(t, \vec{z}) =: \sum_{j=1}^{q_i} \pm h_{ij}(t, \vec{z})$$

wherein all $h_{ij}(t, \vec{z})$ are positive functions. We recommend to note these on a sheet.

We finally aim to detect specific summands that match with transition, death and creation processes. Especially the detection of transition processes is of prior interest, while birth and death processes usually result from the remaining parts of the system. Therefore, it might sometimes be required/useful to transform specific elements of the equation system, but caution is required when dividing by system variables $\vec{z}_i$ as these might potentially become zero in the microscopic model.

For all $i \in \{1, \dots n\}$ and $j \in \{1, \dots, q_i\}$ for which $h_{i,j}(t, \vec{z})$ has a **negative** sign:

Is there a second summand $h_{i',j'}$ with positive sign left in the right hand side vector with $$h_{i',j'}(t, \vec{z}) = h_{i,j}(t, \vec{z})$$ and $i \neq i'$?	
Yes	Continue below
No	If all summands have been investigated: Continue with Question 4.31.

Can $\vec{z}_i$ be dragged out of $h_{i,j}(t, \vec{z})$ without leading to irregularities (like dividing through zero)?	
Yes	In case $\omega(t, i, i', \vec{z})$ has not been set, set $\omega(t, i, i', \vec{z}) := \frac{h_{i,j}(t,\vec{z})}{\vec{z}_i}$. Otherwise add the stated quotient to the existing rate. Strike both summands $h_{i,j}(t, \vec{z})$ and $h_{i',j'}(t, \vec{z})$ from the right-hand side, as they are already dealt with.
No	If all summands have been investigated: Continue with Question 4.31.

Question 4.31:

Now, we aim to find possible creation-processes in the equation system. For all remaining summands $h_{i,j}(t, \vec{z})$ with a positive sign:

Is there an index $i' \in \{1, \ldots, n\}$ so that $\vec{z}_{i'}$ can be dragged out of $h_{i,j}(t, \vec{z})$ without leading to irregularities (like dividing through zero)?	
Yes	In case $c(t, i', i, \vec{z})$ has not been set, set $c(t, i', i, \vec{z}) := \frac{h_{i,j}(t,\vec{z})}{\vec{z}_{i'}}$. Otherwise add the stated quotient to the existing rate. Strike the summand $h_{i,j}(t, \vec{z})$ from the right-hand side, as it is already dealt with. If all summands have been investigated: Continue with Question 4.32.
No	In case $C(t, i, \vec{z})$ has not been set, set $C(t, i, \vec{z}) := h_{i,j}(t, \vec{z})$. Otherwise add the stated number to the existing rate. Remove the summand $h_{i,j}(t, \vec{z})$ from the right-hand side, as it is already dealt with. If all summands have been investigated: Continue with Question 4.32.

Question 4.32:

Now we aim to find possible death-processes in the equation system. For all remaining summands $h_{i,j}(t, \vec{z})$ (they all need to have a negative sign now!):

Can $\vec{z}_i$ be dragged out of $h_{i,j}(t, \vec{z})$ without leading to irregularities (like dividing through zero)?	
Yes	In case $d(t, i, \vec{z})$ has not been set, set $d(t, i, \vec{z}) := \frac{h_{i,j}(t,\vec{z})}{\vec{z}_i}$. Otherwise add the stated quotient to the existing rate. Strike the summand $h_{i,j}(t, \vec{z})$ from the right-hand side, as it is already dealt with. If all summands have been investigated: Continue with Question 4.33.
No	This unfortunately leads to a problem. The system indicates a death rate that does not consider if enough sub-systems are available for removal. One may attempt to assign a universal death rate $D(i, t, \vec{z}) := h_{i,j}(t, \vec{z})$ that a sub-system in state s_i is removed by the overall system, but we cannot give any guarantee about the fit of the microscopic model with respect to the macroscopic differential equation system. In case a solution is found, remove the summand $h_{i,j}(t, \vec{z})$ from the right-hand side, as it is already dealt with. If all summands have been investigated: Continue with Question 4.33.

Question 4.33:

If successful, the right-hand side of $\vec{h}$ should be empty and all summands have been used as specific rates. The mean-field vector $\vec{X}(t)$ of the microscopic model defined by these rate functions approximates the solution $\vec{z}(t)$ of

$$\vec{z}' = \vec{H}(t, \vec{z}) = \beta \vec{F}\left(t, \frac{\vec{y} - \alpha}{\beta}\right).$$

Consequently, the transformed mean-field vector

$$\frac{\vec{X}(t) - \alpha}{\beta}$$

is an approximation of the solution $\vec{x}$ of the given ODE model

$$\vec{x}' = \vec{f}(t, \vec{x}).$$

4.6.5 Scientific Background of the Macro-to-Micro Instruction

In this section we explain the scientific background of the step-by-step instruction defined in the last section 4.6.5. We will conclude this by some thoughts about how this process can be generalised to partial differential equation models, i.e. microscopic models with, at least, one continuous state dimension.

Clearly, the key idea behind the process is to modify the given ODE system, so that it can be identified with the mean-field equation stated in the dynamic population mean-field theorem for discrete state-space Corollary 5.2. Hereby a few observations make the situation more complicated:

- Any mean-field equation according to Corollary 5.2 is solved by a **positive** function vector $\vec{\varphi}(t) \geq 0$. This can be seen investigating $C \equiv 0$. The constant solution $\vec{\varphi}(t) \equiv 0$ can be easily seen to pose for a steady state of the mean-field equation. Hence, a solution with positive initial values cannot become negative. Adding a $C \geq 0$ can only shift the solution towards the positive direction.

- The initial values of $\vec{\varphi}$ of (4.42) in Corollary 5.2 are $\mathcal{O}(N(0))$. If $N(0)$ is small, the mean-field theorem cannot lead to satisfactory results.

- Any rate coefficient that appears on the right hand side of (4.42) in Corollary 5.2 is a function greater than zero, as it results from probabilities.

These three problems are overcome as follows: First the equation is translated, so that the solution vector of the given ODE system is always positive. Keep in mind that for a finite timespan $[0, t_{\text{end}}]$ this is always possible, but not always useful. In case such a shift via α was necessary to obtain positive solution functions, it follows that the given ODE system itself is not capable for being a mean-field model of a microscopic model, yet a shifted (translated) version of it might be.

The second parameter β can be reasoned by considering the initial values of the model. While the fist transformation guarantees positivity of the solutions, the second one, a scaling factor, guarantees a sufficiently large $N(0)$ to keep the differences between the microscopic model and the mean-field model results small. Again, consider that in case such scaling is necessary, the ODE model itself is basically not capable for inverse mean-field analysis.

To obtain positive rate functions, the summands of the shifted ODE system need to be investigated for positivity. Clean negative contributions to the right-hand-side of the ODE can be identified with death rates or transition rates to different states, clean positive contributions can be identified with birth and transition rates from different states. Contributions which cannot be decided for positive or negative, need to be split into positive- and negative- parts to make an identification possible. The choice to investigate the right-hand-side for transitions before considering death and birth processes is not random. If done in a different order, every transition term would be attributed to a birth and a death process instead. A model without transitions would occur, which is, though correct, not useful.

Finally a few ideas are summarised to extend these ideas to the continuous case. Based on experience so-called transport terms, i.e. terms in the shape

$$a(t, x, \varphi(\vec{t}, x)) \frac{\partial \varphi(\vec{t}, x)}{\partial x}$$

results from processes wherein the continuous state of a sub-model (on the average) increases with velocity a. E.g. for a time-discrete model with step-size Δt a transition probability (density) like

$$\tilde{P}(I(t + \Delta t) = x + y | I(t) = x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y - a\Delta t)^2}{2}}$$

would result in the transport term described earlier. One of the best examples for a process that results only in a transport term is ageing. Hereby the continuous variable (age) steadily increases with speed $a = 1$ and is completely deterministic, i.e.

$$\tilde{P}(I(t + \Delta t) = x + y | I(t) = x) = \delta(y - \Delta t).$$

On the contrary, diffusion terms like

$$b(t, x, \varphi(\vec{t}, x)) \frac{\partial^2 \varphi(\vec{t}, x)}{\partial x^2}$$

are caused by stochastic uncertainty of such a movement. The more uncertain the deterministic variable increases/decreases, the larger the diffusion coefficient is. While the ageing example would lead to a vanishing diffusion, the Gaussian distribution in the first example would result in a constant diffusion parameter ($b = 1/2$). Finally, any terms with higher order would then depict the stochastic uncertainty of the previous order. As stochastically fluctuating diffusion process, e.g. a random walk that uses another random walk as variance, would result in terms of third order etcetera.

CHAPTER 5

Application of Mean-Field Theorems – Case Studies

Equipped with the Theorems for mean-field analysis and the stochastic basics introduced in the last two chapters we are given a powerful tool for aggregated analysis of almost any microscopic model. Especially the presented decision trees in Sections 4.6.4 and 4.6.2 can straight forwardly be used to determine suitable mean-field models for microscopic models and vice versa.

Yet, the devil is in the details when applying the step-by-step recipes. Instructions like *"Investigate the behaviour of the agents and summarise all possible states"*, *"Make a list of all possible transitions from one state to another"*, or *"For $t \in T$ calculate the corresponding transition probabilities $P(I(t+\Delta t) = s_j | I(t) = s_i, \dots)$"*, might seem straight forward at first, but may be very hard and confusing to apply. Occurring problems usually have nothing to do with the application of a mean-field theorem, as it is essentially applying a formula. Yet, the process how the microscopic model is best formalised quickly turns out to be almost an art-form as it requires a lot of creativity, instinct, experience and deep knowledge about the stochastic methods in Chapter 3. If important parts of the model are left out or probabilities are calculated falsely, the mean-field approximation will not be a suitable approximation for the microscopic model. Vice versa, if calculated probabilities are used wrongly in the set-up, a microscopic model will not suit to the macroscopic model as well.

Thus, we aim to state a couple of case studies in this chapter where the theorems and the step-by-step instructions are applied on a variety of examples. Hereby we

a) want to give the reader an idea how specific probabilistic concepts can be applied to find valid mean-field models, receiving as little approximation-error as possible,

b) want to underline different aspects of microscopic models with respect to their mean-field behaviour,

c) want to give examples how mean-field analysis can be used for answering modelling and simulation questions, and finally

d) want to motivate the final classification in the last chapter of this thesis.

Some of these aspects focus on the difference between continuous and discrete, some of them should give an example how different time-update influences the model outcome. Some others will show the limits of mean-field analysis and that some models simply cannot be analysed on the macroscopic level using this technique.

5.1 Different Time Concepts in Microscopic Models

Before actually performing mean-field analysis we investigate the differences between different time-update methods in microscopic models in detail to validate the ideas mentioned in Section 4.6.3. As mentioned several times before there are in principle two possibilities how time is updated in microscopic models:

- the model is updated in (equidistant) discrete time-steps, or
- the model is updated by state-events.

We will occasionally denote the latter as time-continuous update of the model, as the underlying time-space is a continuous one.

It is obvious, that both concepts have advantages and disadvantages: Discrete-time update, on the one hand, classically makes the model easier to parametrise. Also, communication between model parts (e.g. via contacts) can be implemented more straight forward. Moreover, it can sometimes be easier to verify as one may directly compare the model before and after a time step to check if the model did what it was supposed to. On the other hand, time-continuous update makes it easier to maintain the correct causal chronology of events in the model. This becomes clear thinking of discretely updated models with longer time-steps: Suppose a sub-model is scheduled to experience two or more state changes in one time-step, the modeller needs to take care which of them should be executed first. Hereby rather artificial rules must be "invented" which have usually nothing to do with modelling.

To benefit from the advantages of both concepts, it is necessary to understand the link that exists between them. The key idea to this link was already mentioned in Section 4.6.3. Understanding the link not only makes it possible to understand why specific things cannot be depicted by either of the two approaches, it is also possible to remodel some existing models that are developed with either of the two concepts.

5.1.1 Time Scaling of Stochastic Models

As it is easier, we will first discuss, what it means to compare two time-discrete models with different step-sizes.

Let us investigate a fictional microscopic model M^1 that is updated in discrete time-steps of length one and suppose that any sub-model (e.g. I^1) with state s is always planned to switch to state s' during the next time-step with a certain probability p_1. I.e.

$$P(I^1(t+1) = s' | I^1(t) = s) = p_1, \quad \forall t \in \{0, 1, 2, \ldots, t_{\text{end}}\}. \tag{5.1}$$

Now let us consider a second model $M^{0.5}$ that updates in equidistant time-steps of length 0.5 and investigate the same problem, now considering sub-model $I^{0.5}$ and its probability to switch to s' in one of the time steps with length 0.5:

$$P\left(I^{0.5}\left(t+\frac{1}{2}\right)=s'|I^{0.5}(t)=s\right)=p_{0.5},\quad \forall t\in\{0,0.5,1,1.5,\ldots,t_{\text{end}}\}. \tag{5.2}$$

Say, we developed M^1 and $M^{0.5}$ for the same purpose or we simply want to rescale M^1 to smaller time-steps, we surely expect the two models to behave alike. Hence, we raise the question:

How do we have to choose $p_{0.5}$ so that $M^{0.5}$ behaves like M^1?

To answer this question, we have to specify what "behaves like" means, in this sense. Clearly, we might be interested in the probability that $I^{0.5}(t+1)=s'$ is equivalent to $I^1(t+1)=s'$. Any two models that fulfil the mentioned property for all agents will be said to fulfil the **Scaled Transition-Probability (STP)** condition henceforth. By the laws of conditional probability, we get

$$\begin{aligned}
&P(I^{0.5}(t+1)=s'|I^{0.5}(t)=s)\\
&\quad=P(I^{0.5}(t+1)=s'|I^{0.5}(t+0.5)=s\wedge I^{0.5}(t)=s)P(I^{0.5}(t+0.5)=s|I^{0.5}(t)=s)\\
&\quad+P(I^{0.5}(t+1)=s'|I^{0.5}(t+0.5)=s'\wedge I^{0.5}(t)=s)P(I^{0.5}(t+0.5)=s'|I^{0.5}(t)=s)\\
&\qquad=P(I^{0.5}(t+0.5)=s'|I^{0.5}(t)=s)P(I^{0.5}(t+0.5)=s|I^{0.5}(t)=s)\\
&\qquad+P(I^{0.5}(t+0.5)=s'|I^{0.5}(t)=s')P(I^{0.5}(t+0.5)=s'|I^{0.5}(t)=s)\\
&\qquad\qquad=p_{0.5}(1-p_{0.5})+(1-p'_{0.5})p_{0.5},
\end{aligned}$$

defining $P(I^{0.5}(t+0.5)=s'|I^{0.5}(t)=s')=:(1-p'_{0.5})$. Thus, we need to solve

$$p_{0.5}(1-p_{0.5})+(1-p'_{0.5})p_{0.5}=p_1 \tag{5.3}$$

in order to find a probability for $p_{0.5}$ so that the STP condition is fulfilled. Easily seen, (5.3) cannot be solved without additional information about other state-changes of the model as $p'_{0.5}$ can be interpreted as the probability that the agent re-switched to a different state in the second time interval, after reaching s' in the first time interval. That means that within two time-steps, model $M^{0.5}$ may switch to a state that might not even reachable with M^1 in the same amount of time. Consequently, it is in general **impossible to perfectly rescale a stochastic model to a different time-step size**, without adding additional rules or restrictions e.g. via a memory.

Nevertheless it is possible to find good approximations in certain special cases. Say the probability that the sub-model leaves state s' within $(t,t+0.5]$ is quite small, we may assume $(1-p'_{0.5})\approx 1$ and find

$$p_{0.5}(1-p_{0.5})+(1-p'_{0.5})p_{0.5}\approx p_{0.5}(1-p_{0.5})+p_{0.5}=2p_{0.5}-p_{0.5}^2=p_1$$

which has the only feasible (in the sense of probabilities) solution

$$p_{0.5}=1-\sqrt{1-p_1}=1-(1-p_1)^{0.5}. \tag{5.4}$$

This formula can easily be extended to any other time-step size δ leading to

$$p_\delta = 1 - \sqrt{1 - p_1} = 1 - (1 - p_1)^\delta. \tag{5.5}$$

This formula is one of the two formulas in this section which may be used to validly rescale probabilities to different time-steps. Although this concept really works well to satisfy the STP condition, a completely different measure of quality heavily fails: The **Expected Time to State-change (ETS)** from s to s'

$$\mathbb{E}(i : I(t+i) = s' \wedge \forall 1 < j < i : I(t+j) = s | I(t) = s). \tag{5.6}$$

Clearly, we expect both models to result in equivalent time expectations and we will state that two models fulfil the ETS condition if the expected state-change time instant is equal in both models. For model M^1 the expected time to state-change is geometrically distributed: For any positive integer i, the probability that i is the first time-step for which a sub-model I^1 switches to s' can be calculated via

$$P(\text{switched in time-step } i-1 \to i) \prod_{j=1}^{i-1} P(\text{did not switch in time-step } j-1 \to j)$$
$$= p_1 \prod_{j=1}^{i-1} (1 - p_1) = p_1 (1 - p_1)^{i-1}.$$

Consequently, the ETS is given by

$$\mathbb{E}(i : I^1(t+i) = s' \wedge \forall 1 < j < i : I^1(t+j) = s | I^1(t) = s) = \frac{1}{p_1}. \tag{5.7}$$

For hints how the expected value of the geometric distribution can be calculated to the expression above, the reader is referred to standard literature about probability theory. For $M^{0.5}$ the situation is slightly different. Here it is appropriate to first investigate the expected amount of time-**steps** to state-change

$$\mathbb{E}(i : I^{0.5}(t+0.5i) = s' \wedge \forall 1 < j < i : I^{0.5}(t+0.5j) = s | I^{0.5}(t) = s). \tag{5.8}$$

Here we again find a geometric distribution and $\mathbb{E}(\dots) = \frac{1}{p_{0.5}}$ results. As *time to state-change* is equivalent to *number of time-steps until state-change* multiplied by 0.5 also

$$\mathbb{E}(i : I^{0.5}(t+i) = s' \wedge \forall 1 < j < i : I^{0.5}(t+j) = s | I^{0.5}(t) = s) = 0.5 \frac{1}{p_{0.5}}. \tag{5.9}$$

Consequently, $p_{0.5}$ needs to be calculated according to formula

$$p_{0.5} = p_1 \cdot 0.5 \text{ , or, generalised to steps of length } \delta,\ p_\delta = p_1 \cdot \delta \tag{5.10}$$

to guarantee that M^2 fulfils the ETS condition. This is the second derived formula in this section which can be used to rescale probabilities. Easily seen, both formulas lead to different results.

Concluding, both scaling formulas (5.5) and (5.10) can be used to create a model M^δ that behaves similar to M^1. As the one fulfils the ETS condition while it fails to fulfil the STP condition and the other vice versa, neither of them are precisely correct and a scaled model M^δ can never be expected to produce stochastically equivalent results.

Yet, interestingly, both formulas lead to similar results when probability p_1 is small. This can be seen in Figure 5.1 and can be reasoned by Taylor approximation

$$1 - \sqrt{1 - p_1} = 1 - (1 - \frac{1}{2}p_1 + \mathcal{O}(p_1^2)) = \frac{1}{2}p_1 + \mathcal{O}(p_1^2).$$

Clearly, this idea can be generalised to arbitrary step-size δ as well

$$1 - (1 - p_1)^\delta = 1 - (1 - \delta p_1 + \mathcal{O}(p_1^2)) = \delta p_1 + \mathcal{O}(p_1^2). \tag{5.11}$$

Hence, the smaller the original probability p_1, the smaller the difference between the two versions of scaled probabilities $p_{0.5}$ (or p_δ) and, correspondingly, the smaller the difference between STP and ETS. Although these two measures for quality of the scaling are not sufficient to achieve stochastic equivalence of the two models, their conservation is yet crucial for the success of the scaling. Summarising

Time scaling from time-step length 1 to $\delta < 1$ works well if the involved probability p_1 is sufficiently small. Hereby both formulas (5.5) and (5.10) can be applied as they lead to similar results.

As (5.10) is clearly the simpler one, **we will use this henceforth**. Figure 5.1 shows a direct comparison of a model with transition probability p_1 and update with time-step size 1 and a corresponding model with $p_{0.5} := 0.5p_1$ and time-step size 0.5. A total of 10000 sub-models were simulated starting in state s. They were observed until they finally reached state s'. The figure shows histograms of the simulation results. The upper left plot showing model M^1 with $p_1 = 1$, clearly indicates that all 10000 sub-models switched to state s' instantaneously after the first step. The corresponding plot with model $M^{0.5}$ and probability $p_{0.5} = 0.5$ is found directly below. Here only about 7500 of 10000 sub-models switched in model $M^{0.5}$ until $t = 1$ (i.e. during the first two steps of the model) which results in an empirically calculated STP of about 0.75. Hence, the resulting STPs differ by a large margin (almost 0.25). Nevertheless, the empirical ETS, equivalent with the mean of the observed transition times, is almost equivalent due to the choice of the scaling formula (Note, that many sub-models already switched to s' during the first time-step of length 0.5, which are not visible in the histogram due to the choice of the bins). The equivalence of the empiric ETS can be seen in the other two comparisons as well, all displaying the results for M^1 on top and the corresponding results for $M^{0.5}$ below. Choosing a smaller value for p_1, also the empiric STPs of both models are getting closer. Finally, for $p_1 = 0.2$ both models already generate almost undistinguishable histogram curves.

5.1.2 Discrete to Continuous Time

Besides proving that time-scaling from 1 to time-steps $\delta < 1$ works well only for small transition probabilities p_1, Section 5.1.1 also indicated that the important measure **Time to State-change**

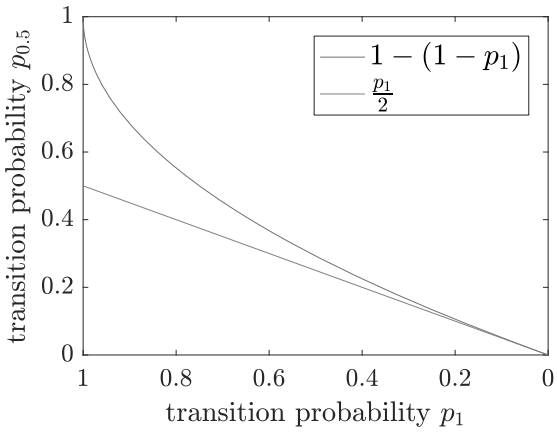


Figure 5.1: Comparison of the two possible formulas for scaling a model from time-step length 1 to time-step length 0.5.

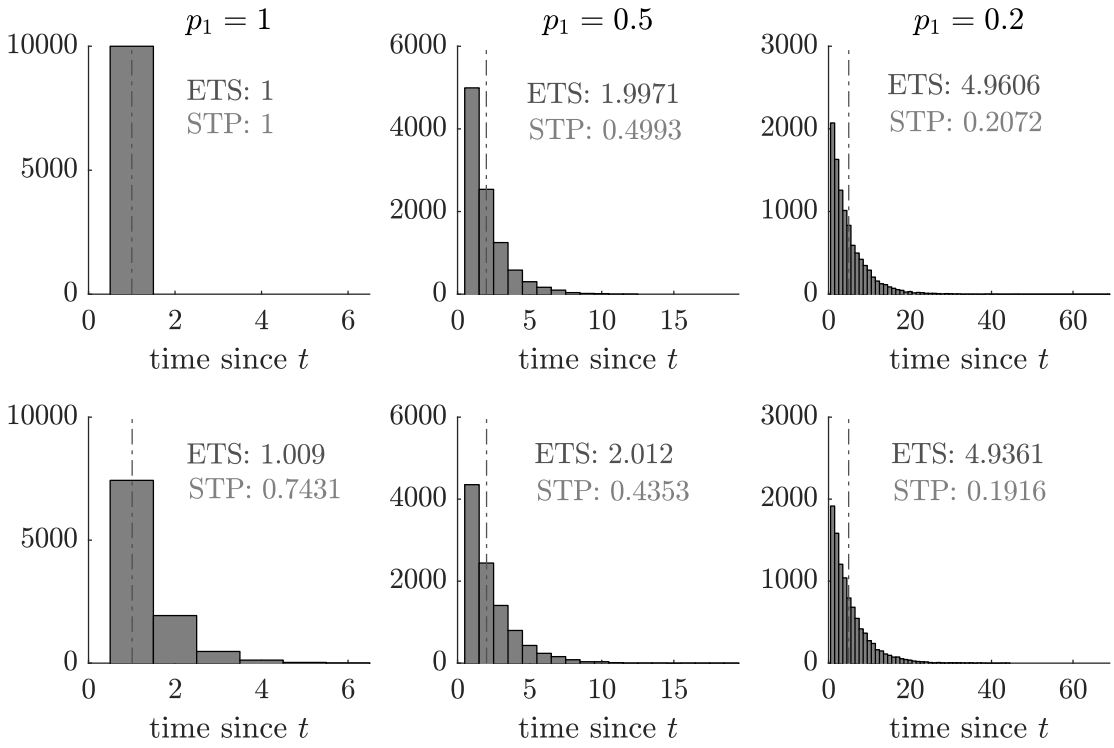


Figure 5.2: Histogram plots of different time-scaling scenarios. The model with step-size 1 is shown in the top three plots, the corresponding model with step-size 0.5 is shown below. Formula (5.10) was used to calculate $p_{0.5}$ from p_1.

(TS) does not only have mean $\frac{1}{p_\delta}$, but is geometrically distributed – independent of how small the step-size δ is chosen. Hence, we may ask: what happens with the limit $\delta \to 0$? To answer this question, we can make use of an important relationship between two distributions:

Corollary 1.1: Convergence of Geometric to Exponential Distribution

In case a random variable X_n is distributed geometrically with parameter $p_n = \frac{p_1}{n}$, then $\lim_{n\to\infty} \frac{1}{n} X_n$ exists and is exponentially distributed with parameter p_1.

Proof. For any $a \in \mathbb{R}$ we calculate

$$P\left(\frac{1}{n}X_n > a\right) = P(X_n > na) = (1-p_n)^{\lfloor an \rfloor} = \left(1 - \frac{p_1}{n}\right)^{\lfloor an \rfloor}$$
$$= \left(1 - \frac{p_1}{n}\right)^{an} \left(1 - \frac{p_1}{n}\right)^{\lfloor an \rfloor - an}$$

The second factor clearly converges to 1 as $\lfloor an \rfloor - an$ is bounded. The first factor converges towards $\exp(-p_1)^a = \exp(-ap_1)$. Hence

$$P\left(\frac{1}{n}X > a\right) \underset{n\to\infty}{\longrightarrow} \exp(-ap_1), \tag{5.12}$$

which can be identified as the probability law of the exponential distribution. ■

Hence, when we scale time to infinitesimal small time-units – we may call this model M^0 – the TS follows an exponential distribution with parameter p_1. Clearly, we are not able to simulate any model with infinitesimal small time-steps, but the clever feature of this concept is that we don't have to.

Knowing that a state-change appears in an exponentially distributed time-instant in the future, we may simply sample an exponentially distributed random variable with parameter p_1 – here, usually called **rate-parameter** – and jump to this point in time. We changed the paradigm:

- **Positive step-size** $M_\delta, \delta > 0$: Probability **that an event happens** in a given time-span.

$$\downarrow$$

- **Infinitesimal small step-size** M_0: Distribution of the time **when the event happens**.

Hereby we notice the concept of event-driven simulation as stated in 2.1.5 and the idea of (simulation of) time-continuous stochastic processes explained in 3.4.4. Figure 5.3 shows a direct comparison of M^1 and M^0 for $p_1 = 0.2$. Hereby the first model was simulated tracking 10000 sub-models for as many time-steps as needed until a state-change occurred and the second model was executed simply generating random samples from an exponential distribution. The resulting histograms are almost equivalent.

In this and the preceding sections we presented a concept that makes it possible to a) rescale microscopic models to different (smaller) time-steps and b) switch from a time-step-driven to an

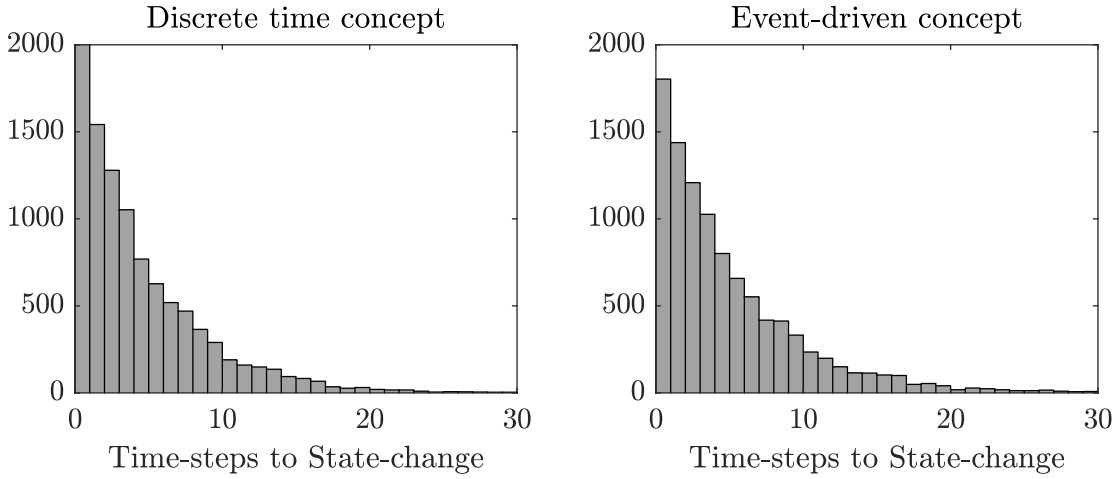


Figure 5.3: Histogram plots comparing discrete-time model M_1 and continuous-time model M_0. Transition probability of $p_1 = 0.2$ was used for the discrete-time model, which was used as rate-parameter in the exponential distribution for the continuous-time model.

event-driven model. In case the involved transition probabilities are sufficiently small the model outcomes of the scaled model and the event-driven model will behave like the unscaled model. Based on experiments, we deem 0.2 for a rule-of-thumb bound for the transition probability, above which scaling cannot be applied without additional thoughts.

5.1.3 Scaling Case Study: Microscopic Population Model

To validate the partly theoretical, partly experimental findings of the last section we chose to apply the scaling concepts on a simplified, academical model for the population of Austria. Hereby I'd like to offer my personal thanks to my colleague Melanie Zechmeister who is heavily responsible for the success of this case study [Zechmeister, 2017].

Although it might also be interpreted as an agent-based model, we prefer to treat it as microsimulation model for its simplicity.

Model 1.1: Microsimulation Model for Austria - Discrete Time

The investigated discrete-time microsimulation population model for predictive simulation of the population of Austria is organised as follows:
Initial Setup:

- A total number of $N \approx 8000000$ individuals, henceforth denoted as agents are generated, each assigned a unique ID.

The model is updated in discrete time-steps of length dt **years**.
Dynamics:

- Each time step, each agent is addressed once. Hereby the following decisions are made:
- The agent dies with probability $dt \cdot \frac{1}{80}$. This number was chosen to ensure a life expectancy of about 80 years.
- In case it did not die, the agent creates an offspring with probability $dt \cdot 0.0096$. This number was derived from the average number of children per 1000 inhabitants in Austria per year according to the Austrian Bureau of Statistics.
- Now, an agent that did not die before, may emigrate with probability $dt \cdot 0.0115$. Also, this number was determined to be as realistic as possible using freely available datasets from to the Austrian Bureau of Statistics. As emigrated agents cannot be considered after their emigration date, they are treated like dead agents.
- Finally, at the end of each time-step, the population is additionally increased by 2% in order to model immigrants. Hence, $dt \cdot$ population size $\cdot\, 0.02$ new agents are created from scratch and added to the agent population.

The present model would in general allow rescaling, but slight errors need to be expected. A reason for this is the process of immigration (probabilities for death, birth and emigration are sufficiently small for almost all $dt < 1$). This process is deterministic and has a huge influence on the model outcome. Here a different rescaling-mechanism is required

The following event-based model was derived based on the ideas presented in the last section.

Model 1.2: Microsimulation Model for Austria - Event-Based

The investigated event-based microsimulation population model for predictive simulation of the population of Austria is organised as follows.
Initial Setup:

- A total number of $N \approx 8000000$ individuals, henceforth denoted as agents are generated, each assigned a unique ID.
- For each agent, the following numbers are sampled: An exponentially distributed random death-date $D \sim Exp(1/80)$, an exponentially distributed random emigration-date $E \sim Exp(0.0115)$ and an exponentially distributed random offspring-date $O \sim Exp(0.0096)$. The larger of the emigration and death dates is removed. If the offspring date is larger than the smaller of the emigration or death date, it is removed as well. Remaining dates are added to a so-called event-list.

The model is updated, jumping to the next recent event in the event list. **Dynamics:**

- In case of a death- or emigration-date of an agent, the agent is removed. In case it was an offspring-date, a new agent is created. In this case, new death, emigration and offspring dates need to be sampled for the new-born agent analogously to the initialisation of the model.
- Additionally, the formula $(t_{\text{new}} - t_{\text{old}})\text{population size} \cdot 0.02$ is used to create a number of immigrants that entered the model during the time between the last and the current event. For each of those, new death-, emigration- and offspring-dates need to be sampled analogously to the initialisation of the model.

The results of the comparison were astonishing and surprising. First of all, the discrete-time model was run for different time-step widths from $dt = 0.02$ (i.e. almost weekly steps) up to $dt = 10$ years. Hereby apart from the expected differences between the mean-value curve, mainly two observations could be made:

- The higher the step-width (i.e. the less steps required to reach the end-time) the higher fluctuations could be seen comparing the simulation results. Hence, using a higher number of steps has a smoothing effect on the simulation outcome. This seems to be unintuitive at first: One would probably assume, that any additional source for randomness (as given when increasing the number or model steps) would increase the fluctuations, but it is complementary due to ergodic effects.
- The computational efforts required for the simulation do **not** grow/decay linearly when decreasing/increasing the step-size. I.e. if the step-size is halved the computational error is not increasing by factor 2 but by a smaller one. This might, again, seem unintuitive as halving the step-size leads to the doubled number of time-steps to deal with. Yet one must keep in mind, that also the probabilities are halved making it less likely that state-changes

really take place. Hence, though the number of loop iterations is doubled, time-consuming agent state-changes are a lot less frequent than before.

Finally, comparing the discrete time approaches with the event-driven one leads to the following observations:

- The event-driven model is hardly fluctuating. The fluctuations can be compared with the discrete-time model run with $dt = 0.1$.
- Agent-based discrete event simulation is **not** (at least not always) a massive overhead, neither with respect to the model definition – in this case it was even shorter than the time-discrete one – nor with respect to computation time. The run-time of the event-driven model can be compared with the run-time of the discrete-time model with steps of about 0.2 years – i.e. about two months. Any larger step-size results in a faster executing discrete-time model, any smaller, in a slower one.

Figure 5.4 shows a comparison of the event-driven with the time-discrete model for different time-step widths. For $dt = 2$ and $dt = 0.5$ the resulting sample standard deviations of the time-discrete models are a lot higher than for the event-driven case. Also, the mean value curve slightly differs from the results of the event-driven model. For $dt = 0.1$ and 0.02 the sample mean levels are almost equal, yet the standard deviation can still be seen to shrink for the smaller step-size.

Clearly, the model is not validated as the population development has hardly anything to do with the real population development in Austria. Yet, the case study underlines that it might be worth thinking about rephrasing a microscopic model in a discrete event sense. The event-graph formalism introduced in Section 2.1.5 gives an additional opportunity to uniquely describe the model in a graphical way (see Figure 5.5). The population model is specifically well suited for this process as state events of the agents are very rare (the correspondent probabilities are quite small). In case agents change their state more frequently the computational overhead will surely grow. If, moreover, transition probabilities depend not only on the state of the agent alone, but also on states of other agents or, even worse, on time, it becomes a lot more difficult to rephrase the model in a discrete event sense. I.e. if the death rates would change every year, the event-driven approach would result in a more complex model. For some additional ideas about comparison of discrete-time vs. event-driven (agent-based) models the reader is referred to [Buss and Al Rowaei, 2010].

5.2 Case Studies for the Step-By-Step Processes

Before getting into more details about system-theoretic analysis of the features and difficulties of the link between micro and macro via mean-field analysis, we state two simple case studies that should show how the two step-by-step introductions can be applied in general.

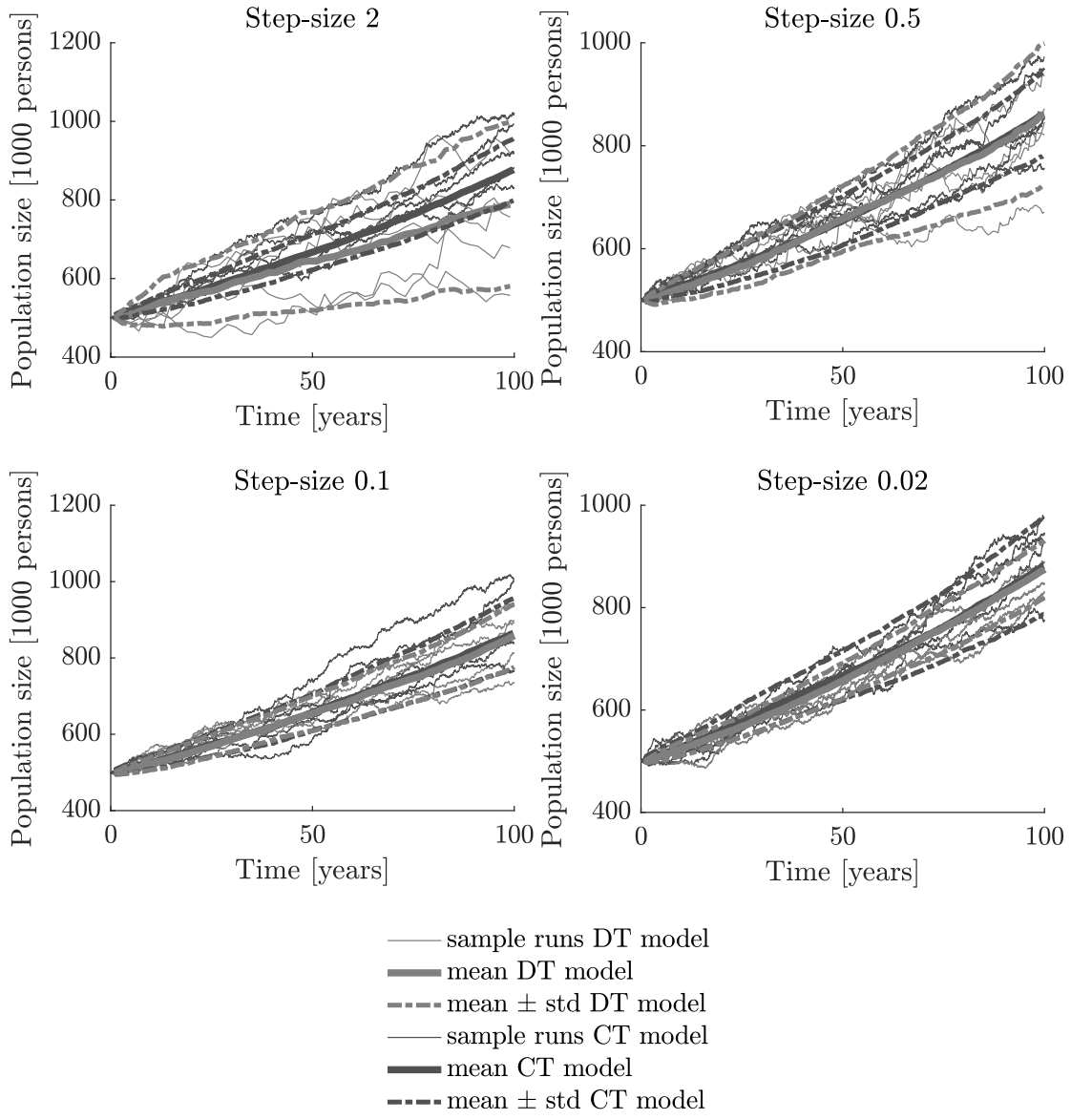


Figure 5.4: The pictures show a comparison between the described event-driven (blue) and discrete time-step (red) population model with different step-sizes dt from 0.02 to 2. The broad, solid lines mark the sample mean of 100 simulation-runs. The dashed lines mark curves which are, each, one sample standard deviation away from the mean. The thin solid lines, finally, show some sample paths of the model.

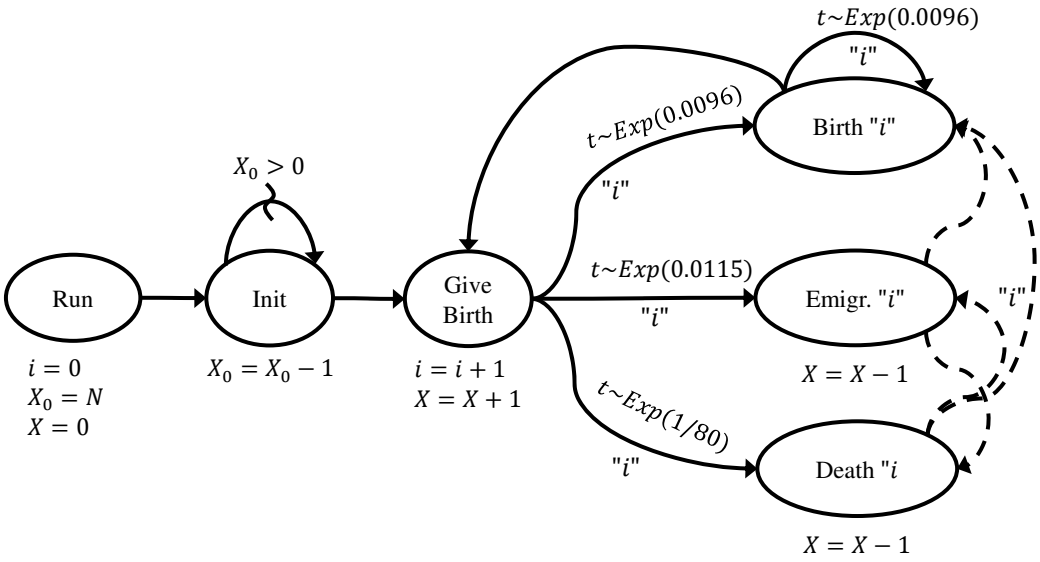


Figure 5.5: Alternative representation possibility of the event-driven population model via the Event Graph. The three event-nodes at the right hand side are initialised with the index parameter i. Hence, they only cancel and reschedule themselves according to this parameter.

5.2.1 Micro-to-Macro Process Case Study: SIR Lattice Model

The first model to apply the step-by-step introduction to is a microscopic Susceptible-Infectious-Recovered (SIR) model. Mentioned SIR model has a long history at TU Vienna and the ARGESIM (ARbeitsGEmeinschaft SIMulation) of Prof. Felix Breitenecker, as it plays a key role in the ARGESIM Benchmark C17 [Breitenecker et al., 2004]. In 2004 it was already subject to a scientific comparison between microscopic and macroscopic models. In 2013 mean-field analysis was applied to this model and published in [Bicher and Popper, 2013] and [Bicher, 2013]. Hence, the following mean-field approximation study for this model should not be considered as anything particularly new. Preferably, it should be seen as a perfect first test for the presented step-by-step instruction in Section 4.6.2.

Model 2.1: SIR Lattice Model

This microscopic SIR model, presented in [Bicher and Popper, 2013] or [Breitenecker et al., 2004], is defined as an agent-based model (ABM) wherein each agent inhabits a site on a lattice. In order to understand the described processes the reader might take a look on Figure 5.6.

Initial Setup:

- First of all let Ω be a discrete, rectangular grid with $M := M_x \cdot M_y$ sites c_{ij}.

- Each site $c_{ij}, i \in \{1,\dots,M_x\}, j \in \{1,\dots,M_y\}$ itself is partitioned into four site-fractions $\{c_{i,j,1,1}, c_{i,j,2,1}, c_{i,j,1,2}, c_{i,j,2,2}\}$.
- Let $N \leq 4 \cdot M$ be a fixed number of agents.
- At each point t in time each agent $I_n,\ n = 1,\dots,N$ has a certain state $I_n(t)_1 \in \{1,2,3\}$ ($\equiv$ {susceptible, infectious, recovered}) and a certain position on the grid

  $$I_n(t)_2 \in \{(i,j,k,l) :\ i \in \{1,\dots,M_x\}, j \in \{1,\dots,M_y\}, k,l \in \{1,2\}\}.$$

  To simplify the speech, the state of the agent is usually called as a prefix of the word agent: e.g. "infected agent".
- Initially a number of N_1, N_2, N_3 with $N_1 + N_2 + N_3 = N$ agents with states $1,2,3$ are randomly distributed on the fractioned grid. I.e. their position is unique (and remains unique).

The simulation is time-discrete with a finite number of time-steps $\{1,\dots,t_{\text{end}}\}$.
Dynamics:

- Each of these steps consists of two phases: an infection-phase and a movement-phase. During each phase each agent is addressed once.
- If, during infection-phase, an infected agent shares the site with a susceptible agent, the susceptible agent's state has a certain probability λ to become infected by its site-mate. Hence, each susceptible agent may get infected by each site-mate that is infectious.
- Moreover, each infectious agent has a certain probability μ to overcome the disease and become immune.
- If all agents are addressed once, movement-phase, necessary for propagation of the disease and inspired by so-called lattice-gas cellular automata, starts. All agents update their position following certain movement rules. Agents with position $(x,y,1,1)$ are moved to $(x-1, y-1, 1, 1)$, agents with position $(x,y,2,1)$ are moved to $(x-1, y+1, 2, 1)$ etcetera. If movement is not possible due to borders, agents switch their position inside the site-fraction. For more detailed information the reader is referred to [Bicher and Popper, 2013] or [Breitenecker et al., 2004].
- After movement-phase, the next time-step starts with infection-phase again.

We aim to perform a mean-field analysis and apply the microscopic to macroscopic instruction from Section 4.6.2. Our target is a mean-field model described by ordinary differential equations (ODEs).

2.1 We start by identifying the microscopic sub-models of our model. Though this seems trivial, we already need to make a first decision: should we consider the site-fraction as

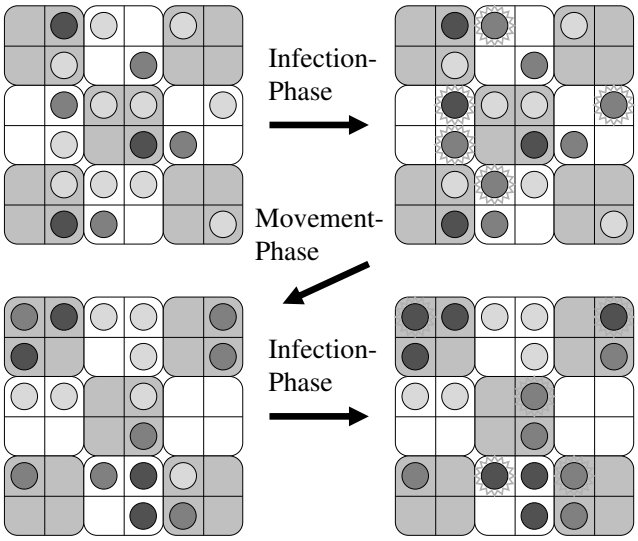


Figure 5.6: Diagram of the time evolution in the microscopic SIR model. This diagram has already been published in [Bicher and Popper, 2013].

sub-models (as it might be the most convenient case in cellular automata) or should we consider the described agents? In fact, both choices would be capable for the mean-field analysis, though clearly the second option is the more obvious and easier one in this case. Hence, N is already the correct name for the variable. As epidemiological models with $N < 50$ hardly make sense, we may continue with 2.2.

2.2-2.5 Here we have to make the next decision. To be precise each agent's state-space basically consists of the tuple

$$(s, i, j, k, l), \quad s \in \{1, 2, 3\},\ i \in \{1, \dots, M_x\}, j\ \in \{1, \dots, M_y\}, k, l \in \{1, 2\}.$$

Hence, clearly the state-space is discrete and would be larger than the total number of agents. As the size of a discrete state-space corresponds to the number of resulting differential equations, this might lead to enormous, unnecessary overhead. For this reason, we define

$$\Gamma := \{1, 2, 3\} \tag{5.13}$$

and try to ignore the position of the agent at first. It will turn out when calculating the transition probabilities if this simplification was legit or not. Alternatively, one could also try to approximate the space coordinates by a continuous variable in the $M_x \times M_y$ rectangle and continue with the continuous version of the Theorem.

Based on the defined state-space we investigate the possible state-changes of an agent. We observe that a change between states 1 and 2 and between 2 and 3 are possible as susceptible agents might get infectious which themselves might recover. Birth and death processes are not present as the model is population-static. Our state-transition list $\{1 \to 2, 2 \to 3\}$ results and we continue with 2.6.

2.6 Clearly, the model is updated in equidistant time-steps with length 1 which makes the remainder of the process considerably easier. We continue with 2.7.

2.7 As $\Gamma = \{1, 2, 3\}$ we note three left-hand sides of differential equations

$$\frac{d}{dt}\phi_i(t) = \ldots, \; i \in \{1, 2, 3\}$$

and continue with 2.8.

2.8 For state 1 the transition $1 \to 2$ can be found in the transition list. Hence, we investigate the probability that an agent switches from 1 to 2 in one time-step. Therefore, we look in the microscopic model definition for the causes for the regarded state-transition 1 to 2 and continue with question in 2.9.

2.9 Clearly, a susceptible agent may only switch to the infectious state if it shares a site with at least one infectious agent. Suppose the agents are approximately uniformly distributed on the grid, then the probability for finding one of the infectious agents on one of the neighboured sites can be calculated with Laplace-space arguments. Note, that this might be the crucial cause for a failure of the mean-field analysis due to ignoring the space-dimension of the model earlier. With ϕ_2 denoting the total numbers of infectious agents at the regarded time instant, then $\frac{\phi_2}{4M}$ describes the probability of finding an infectious agent in a specific site-fraction. As there are 3 neighboured site-fractions the susceptible agent may become infectious by each of them. Enumerating the three neighboured fractions from 1 to 3, let A_1, A_2, A_3 describe the events that the susceptible site became infected by the corresponding site-fraction. As mentioned

$$P(A_i) = \lambda \frac{\phi_2(t)}{4M}, \; i \in \{1, 2, 3\}$$

for all three of them. By the Laws of De Morgan we obtain

$$\begin{aligned} P(A_1 \vee A_2 \vee A_3) &= 1 - P((A_1 \vee A_2 \vee A_3)^c) = 1 - P((A_1^c \wedge A_2^c \wedge A_3^c)) \\ &= 1 - P(A_1^c)P(A_2^c)P(A_3^c) = 1 - \left(1 - \frac{\lambda\phi_2(t)}{4M}\right)^3. \end{aligned}$$

Hence,

$$\omega(t, 1, 2, \vec{\phi}) := 1 - \left(1 - \frac{\lambda\phi_2(t)}{4M}\right)^3 \tag{5.14}$$

can be determined. We moreover subtract $\phi_1(t)\omega(t, 1, 2, \vec{\phi})$ from the first equation and add it to the second one.

2.10 For state 2 we observe that the transition $2 \to 3$ is in the transition list.

2.11 Clearly, an agent in state 2 switches to state 3 with probability μ. Hence

$$\omega(t, 2, 3, \vec{\phi}) := \mu. \tag{5.15}$$

Analogously to the first transition we subtract it, multiplied by $\phi_2(t)$, from the second equation and add it to the third one. As no other state transitions are possible we continue with 2.16.

2.16 The total numbers of agents in either of the states is deterministic and can be set directly:

$$\phi_1(0) := N_1,\ \phi_2(0) := N_2,\ \phi_3(0) := N_3. \tag{5.16}$$

The mean-field model is fully defined and should match the microscopic model, in case the assumption done in the first 2.8. question was correct.

Finally the non-linear differential equation model

$$\frac{d}{dt}\phi_1 = -\phi_1 \left(1 - \left(1 - \frac{\lambda\phi_2}{4M}\right)^3\right) \tag{5.17}$$

$$\frac{d}{dt}\phi_2 = \phi_1 \left(1 - \left(1 - \frac{\lambda\phi_2}{4M}\right)^3\right) - \mu\phi_2 \tag{5.18}$$

$$\frac{d}{dt}\phi_3 = \mu\phi_2 \tag{5.19}$$

is derived which we will henceforth denote as *cubic SIR ODE model* for its cubic terms. Considering the asymptotic equivalence

$$1 - \left(1 - \frac{\lambda\phi_2}{4M}\right)^3 = 1 - \left(1 - 3\frac{\lambda\phi_2}{4M} + \mathcal{O}\left(\left(\frac{\lambda\phi_2}{4M}\right)^2\right)\right) = 3\frac{\lambda\phi_2}{4M} + \mathcal{O}\left(\left(\frac{\lambda\phi_2}{4M}\right)^2\right),$$

we may write $\frac{3\lambda}{4M}\phi_1\phi_2$ instead of the inelegant cubic terms in the first and second equation as long as the density of the infectious agents is small. As mentioned in [Bicher and Popper, 2013] the classic SIR differential equation model by Kermack and McKendrick [Kermack and McKendrick, 1927] results in this case.

Model 2.2: SIR ODE Model [Kermack and McKendrick, 1927]

The classic SIR model by Kermack and McKendrick is given by the following system of three coupled differential equations

$$\frac{d}{dt}\phi_1 = -\frac{3\lambda}{4M}\phi_1\phi_2, \tag{5.20}$$

$$\frac{d}{dt}\phi_2 = \frac{3\lambda}{4M}\phi_1\phi_2 - \mu\phi_2, \tag{5.21}$$

$$\frac{d}{dt}\phi_3 = \mu\phi_2. \tag{5.22}$$

If the fraction $\lambda\phi_2/(4M)$ reaches a level close to 1 during the simulation, the cubic ODE model needs to be used as mean-field approximation, otherwise the classic SIR differential equation model is a good mean-field model as well.

Figure 5.7 shows a direct comparison of solutions of the ABM and (numerically calculated) solutions of the two ODE models. The ABM uses the parameters $\lambda = 0.2$ and $\mu = 0.1$ with $N = 4000$ agents spread on a 40×50 grid. As the plots suggest, the fraction $\frac{\lambda\phi_2}{4M} < \frac{\lambda N}{4M} = \frac{1}{10}$ is small enough that the asymptotic term can be neglected and the classic SIR equations can hardly be seen to differ from the cubic model. The situation is different for $\lambda = 0.8$ and $N = 7000$ agents as seen in Figure 5.8. While both macroscopic models have problems correctly depicting the behaviour of the microscopic model (the probability λ is quite high which is a problem for the link between continuous and discrete time), the cubic model seems to be closer to the microscopic one than the classic SIR model. The fact that in contrast to the cubic ODE model, the classic SIR model depicts the amplitude of the microscopic infectious population seems to be rather a coincidence than a real feature.

Clearly, the most important factor to guarantee the similarity of the microscopic and the derived microscopic models is the movement phase. It guarantees the assumption that the spread of individuals among the grid is approximately uniformly distributed holds and makes it possible to macroscopically model the system without regarding the spatial coordinate. Figure 5.9 shows what happens if the movement phase was skipped using the same parameters as for Figure 5.7. At the beginning of the simulation, the uniform distribution condition is still fulfilled due to the uniformly distributed initial condition of the model, yet, very soon, local effects become visible which lead to incomparably different results.

5.2.2 Variance Theorem Case Study: SIR Lattice Model

Equipped with the transition rates $\omega(t, 1, 2, \vec{\phi})$, $\omega(t, 2, 3, \vec{\phi})$ derived in the course of the last section (all others are $\equiv 0$) we can directly put the discrete variance theorem 5.1 to a test. We will use the simplified version

$$\omega(t, 1, 2, \vec{\phi}) = \frac{3\lambda}{4M}\phi_2, \quad \omega(t, 2, 3, \vec{\phi}) = \mu,$$

as we assume that λ is small enough. Clearly, only the first transition rate depends on the aggregated number. Thus, the nabla operators of the transition rates in the variance theorem

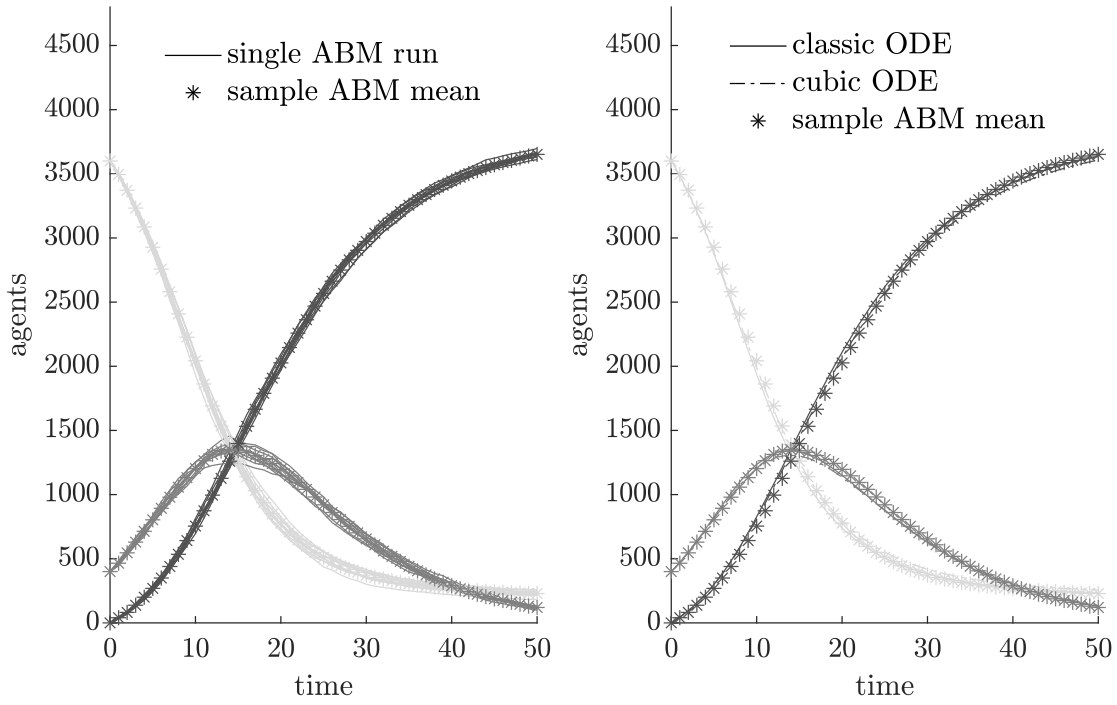


Figure 5.7: Comparison of sample runs of the described SIR ABM and the two derived mean-field models. The left part of the figure shows 10 sample runs of the ABM and their sample mean. The right part shows a comparison of the sample mean with the solution of the classic and the cubic SIR model.

evaluate to

$$\nabla\omega(t,1,2,\vec{\phi}) = \begin{pmatrix} 0 \\ \frac{3\lambda}{4M} \end{pmatrix}, \quad \nabla\omega(t,2,3,\vec{\phi}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{5.23}$$

As we are interested in the temporal behaviour of the variance and $|\Gamma| = 3$ we may apply formula (4.64) which is the specifically modified version of the variance theorem for the variances in a three state model. The following ordinary differential equation system results (with $\overline{\lambda} := \frac{3\lambda}{4M}$ and $\xi_i := \xi_{i,i}$)

$$\frac{d}{dt}\xi_1 = (\phi_1 - 2\xi_1)\phi_2\overline{\lambda} - (\xi_3 - \xi_1 - \xi_2)\phi_1\overline{\lambda} \tag{5.24}$$

$$\frac{d}{dt}\xi_2 = (\phi_1 + \xi_3 - \xi_1 - \xi_2)\phi_2\overline{\lambda} + 2\xi_2\phi_1\overline{\lambda} + (\phi_2 - 2\xi_2)\mu \tag{5.25}$$

$$\frac{d}{dt}\xi_3 = (\phi_2 + \xi_1 - \xi_2 - \xi_3)\mu \tag{5.26}$$

Herein $\phi_i(t),\ i \in \{1,2,3\}$ are the solutions of (5.96)-(5.98). It is practicable to solve (5.96)-(5.98) and (5.24)-(5.26) simultaneously with a numeric ODE solver.

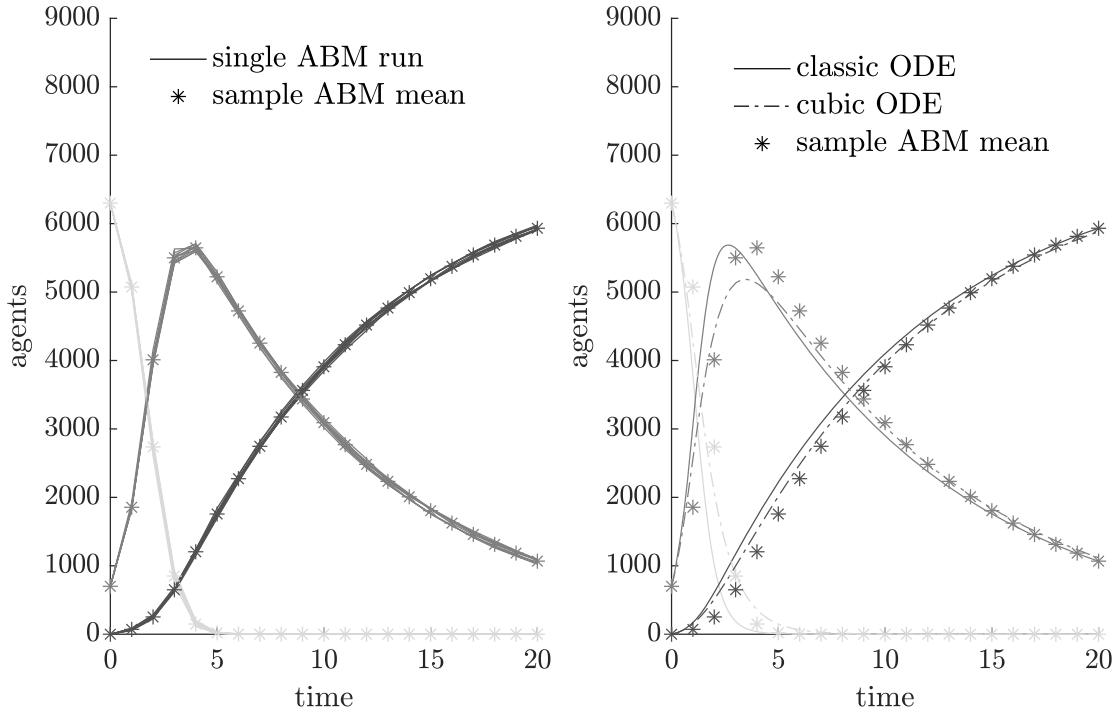


Figure 5.8: Comparison of sample runs of the described SIR ABM and the two derived mean-field models with higher density and infection parameter λ than in Figure 5.7.

One of the most interesting parts of this case study is the choice of the initial conditions of this differential equation system. As stated in the Variance Theorem for Discrete State Space, the initial conditions are given by the initial variances of the initial aggregated numbers. In case the agents are independently and identically distributed random samples according to $P(I(0) = i) = f_i$ as stated in the Theorem, this would result in

$$\begin{aligned}\xi_i(0) = \mathbb{V}(X_i(0)) &= \sum_{k=1}^{N} \mathbb{V}(\mathbb{1}_i(I_k(0))) = \sum_{k=1}^{N} \mathbb{E}(\mathbb{1}_i(I_k(0))^2) - \mathbb{E}(\mathbb{1}_i(I_k(0)))^2 \\ &= \sum_{k=1}^{N} (\mathbb{E}(\mathbb{1}_i(I_k(0))) - \mathbb{E}(\mathbb{1}_i(I_k(0)))^2) = \mathbb{E}(X_i(0)) - \mathbb{E}(X_i(0))^2 = N(f_i - f_i^2),\end{aligned}$$

as $\mathbb{1}_i(\cdot)^2 = \mathbb{1}_i(\cdot)$. As the SIR model uses a "deterministic sampling" at the beginning – i.e. the initial total numbers of X_1, X_2, X_3 are deterministically given by N_1, N_2, N_3 – the initial condition needs to be different. Hence, $\xi_i(0) = 0$ is the correct initial condition for the differential equation system as there are no fluctuations in the initial aggregated numbers. Note, that this is no issue for the correctness of the actual mean-field approximation, as for $N \to \infty$ the initial standard deviation of the aggregated numbers vanishes anyway.

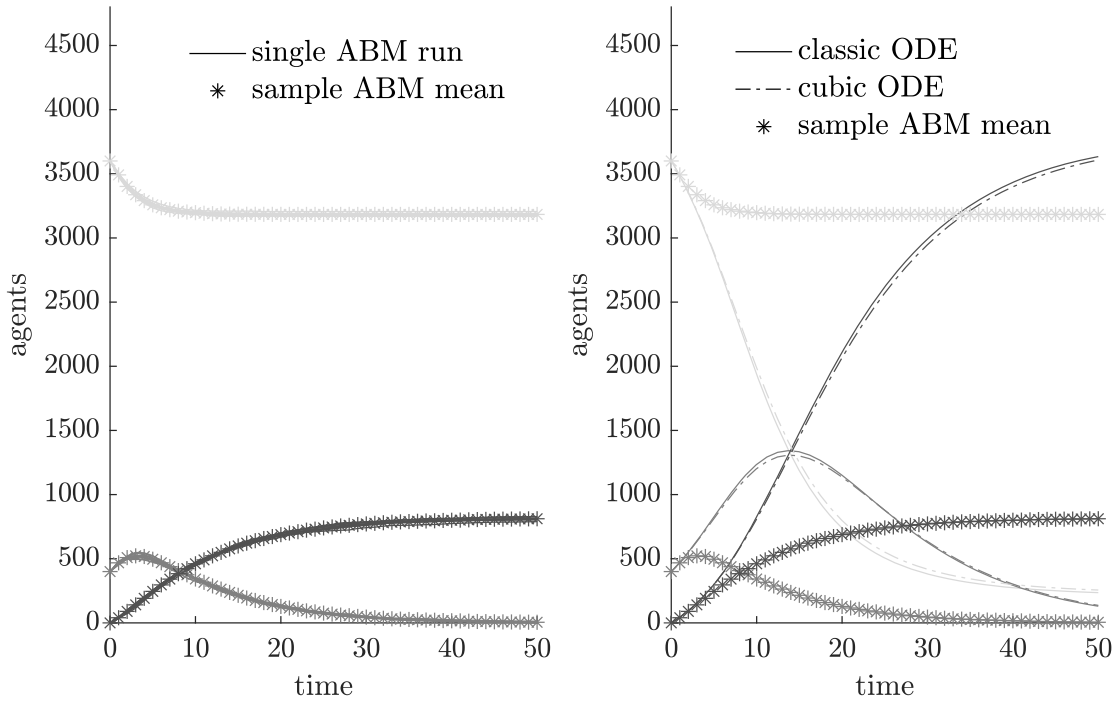


Figure 5.9: Comparison of sample runs of the described SIR ABM and the two derived mean-field models with the same parameters as for Figure 5.7. Yet the movement phase was skipped.

Figure 5.10 shows a comparison of the sample variance of the SIR model and the numeric solution of the stated system of differential equations. It can be seen nicely that the qualitative behaviour is well depicted, yet shortcomings with respect to the quantitative match can be seen. First, these are due to the problem, that the sample variance is a very sensitive measure and a lot of samples had to be drawn to achieve a more-less smooth curve. Secondly, the mentioned issues with initial conditions lead to the slight overestimation of the variance by the solution of the differential equation.

Interpretation of the result curves, in this case, is not particularly interesting. The peak of the variance curve at about $t = 15$ indicates that this particular time is the most sensitive time of the model and that the resulting aggregated values at this point in time should be treated with much more care. Moreover, it is an interesting feature of this model that it loses its fluctuations when time goes to infinity. This feature seems unintuitive at first as the chaos in a model is nothing you would (from a physical interpretation) expect to become smaller with time. Yet, it is clear in this case as the target values converge towards steady states.

5.2.3 Macro-to-Micro Process Case Study: Modified Levins Model

To demonstrate the step-by-step instruction for the inverse mean-field analysis, i.e. how to apply mean-field analysis to obtain a microscopic model out of a given macroscopic one, we decided

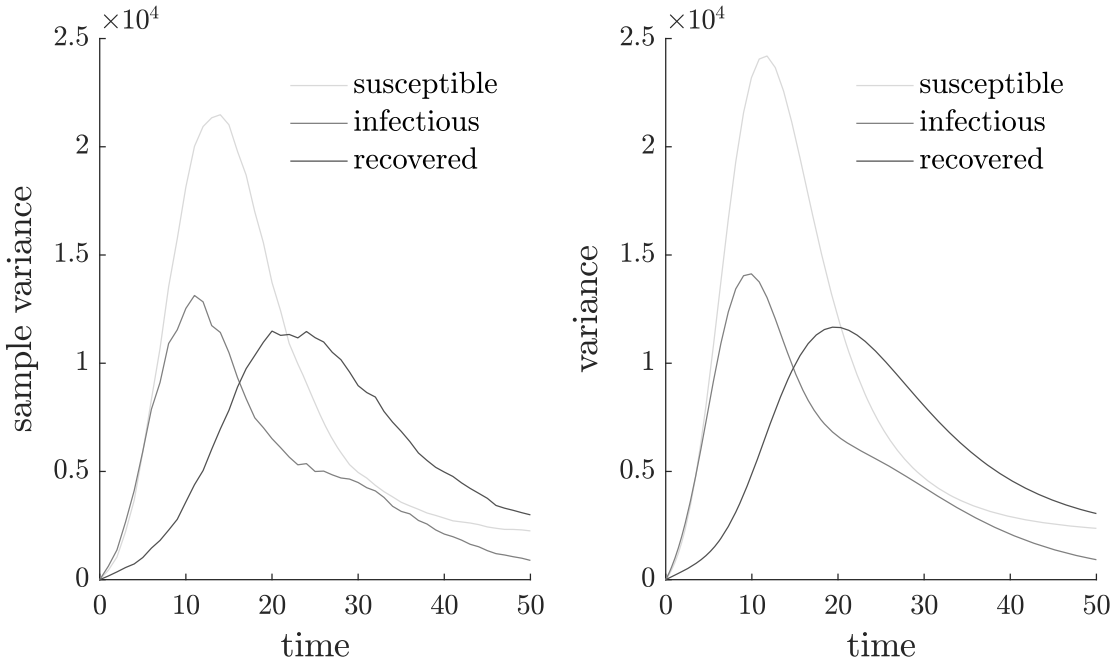


Figure 5.10: Comparison of sample variance of 500 runs of the Lattice SIR model with 100×100 sites (left) and the numerically generated curves of the differential equation system (5.24)-(5.26) (right).

to use a very simple, yet interesting population model given by a scalar, first order differential equation.

Model 2.3: Modified Levins Model (with Immigration)

$$\frac{d\phi}{dt}(t) = \alpha\phi(t)(M - \phi(t)) - \beta\phi(t) + \gamma, \quad \phi(0) = X_0 > 0. \tag{5.27}$$

with parameters $M \gg 0, \alpha > 0, \beta > 0, \gamma > 0$. The latter three may be functions of time, but will be assumed as constant here as it does not make any difference for the step-by-step process.

The main part of the model – i.e. anything apart from the additional constant γ – is commonly known as Levins model, named after its first publisher Richard Levins in 1969 [Levins, 1969]. It can be interpreted as a combination of a logistic growth term that positively influences the growth of a population and a linear mortality term that has a negative impact. The additional constant parameter γ was added to model immigration from outside the model boundaries.

The fact, that Levins formerly established this model to investigate, in his own words, *heterogeneity* and *extinction* seems unusual nowadays: neither can solutions of such differential

equations reach their steady-state 0 (extinction) nor are differential equations particularly well suited for analysis of heterogeneity. Clearly, this is a matter of a different terminology and a different focus. Nevertheless, we might, now, be interested in developing a model that is capable of real extinction and real heterogeneity. Therefore, we want to bring the ideas of the differential equation to the microscopic scope.

We follow the step-by-step instruction from Section 4.6.4:

4.1 The ODE is already given in explicit form and is seen to be of first order. We define $\Gamma := \{s_1\}$.

4.2 As the right hand side of (5.27) fulfils a Lipschitz condition a unique solution of the ODE exists and is steadily differentiable (according to the Theorem of Picard-Lindelöf). Moreover, $\phi \equiv 0$ is a steady state and X_0 is positive. Therefore, $\phi(t)$ is always greater than zero and a shift is not necessary.

4.3 As the Levins model models a population, we assume that X_0 is large enough to be capable to be an initial aggregated number of a microscopic model.

4.4 For positive ϕ we write

$$\frac{d\phi}{dt}(t) = h_1(t,\phi) - h_2(t,\phi) - h_3(t,\phi) + h_4(t,\phi),$$

$$\begin{aligned} h_1(t,\phi) &:= \max(0, \alpha\phi(M-\phi)), \\ h_2(t,\phi) &:= \max(0, \alpha\phi(\phi-M)), \\ h_3(t,\phi) &:= \beta\phi, \text{ and} \\ h_4(t,\phi) &:= \gamma. \end{aligned}$$

Note, that a second index is not necessary as our problem is scalar.

4.5 We deal with the possible state-transitions and observe that h_2 and h_3 have negative signs. As our problem is scalar we don't find them in any other equation.

4.6 Regards birth processes we observe that h_1 and h_4 are positive. For h_1 we may easily drag ϕ out of the equation leading to

$$c(t, s_1, s_1, X) := \frac{\max(0, \alpha X(M-X))}{X} = \max(0, \alpha(M-X)). \tag{5.28}$$

For h_4, dividing by X would be somewhat artificial. Hence, it is probably the more elegant way to set

$$C(t, s_1, X) := \gamma \tag{5.29}$$

and use it as global birth rate.

4.7 We finally consider the death processes. As mentioned, h_2 and h_3 have negative signs. For both also x can be divided easily. Clearly, h_2 leads to

$$\frac{\max(0, \alpha X(X-M))}{X} = \max(0, \alpha(X-M))$$

which adds up with

$$\frac{\beta X}{X} = \beta$$

leading to

$$d(t, s_1, X) := \max(0, \alpha(X-M)) + \beta. \tag{5.30}$$

The microscopic model is in principle defined by its transition rates. Finally creating a specific microscopic model based on the transition/creation/death rates requires some creativity.

Say, we aim to develop an agent-based model that updates in equidistant time-steps. Therefore, we first need to define a small time-step size dt so that

$$c \cdot dt \ll 1,\ C \cdot dt \ll 1,\ d \cdot dt \ll 1.$$

Then we may set

$$P(\text{system creates new agent}) := dt \cdot C := dt \cdot \gamma, \tag{5.31}$$

$$P(\text{agent creates new agent}) := dt \cdot c := dt \cdot \max(0, \alpha(M-X)), \text{and} \tag{5.32}$$

$$P(\text{agent dies}) := dt \cdot d := dt \cdot \beta + dt \cdot \max(0, \alpha(X-M)). \tag{5.33}$$

Though the agent-based model that updates with step-size dt is fully defined by these creation and death probabilities, it is neither a very elegant nor a very picturesque way to define a microscopic model. E.g. the appearance of X in the creation basically implies: To decide if an agent "gives birth" to a new agent, the agent needs to know about the size of the total population X (i.e. the microscopic rule requires the use of macroscopic numbers). To avoid this, we need to find a mechanism, that the macroscopic parameter implicitly occurs. **One option is presented here:** Each agent is assigned a unique integer-valued "position" chosen from the set $\{0, 1, \ldots, M\}$. Moreover, pick a uniformly distributed random spot $s \in \{0, 1, \ldots, M\}$ and assume that X agents are currently alive in the model. Clearly, the probability that the randomly chosen spot is still available is linearly linked to the number of agents that are alive. By the laws of Laplace-Space, it can be calculated by

$$P(s \text{ is available}) = \frac{M-X}{M}.$$

Thus, we establish the following mechanism: To decide about, if an agent generates a new agent, pick a random integer between 1 and M. If the chosen spot is still available, create a new agent with probability $dt \cdot M\alpha$. Moreover, assign the newly generated agent to the chosen spot. Hereby

$$P(\text{agent creates new agent}) = P(\text{spot available}) \cdot dt \cdot M\alpha = dt \cdot \alpha(M-X)$$

Similarly, the second part of the death probability can be implemented: Whenever (by immigration with probability $dt \cdot \gamma$) the maximum number of agents overshoots the maximum

capacity of M – i.e. $X > M$ – each of the agents that cannot be assigned a number between 1 and M is added to a *overshoot basin*. Increasing the dying probability of all agents in the basin to $dt \cdot (\alpha M + \beta)$ also increases the average probability that a random model agent dies: Picking a random agent, we get by the laws of the conditional probability that

$$\begin{aligned} P(\text{die}) &= P(\text{no overshoot}) \cdot dt \cdot \beta + P(\text{overshoot}) \cdot dt \cdot (\alpha M + \beta) \\ &= \frac{M}{X} \cdot dt \cdot \beta + \frac{X - M}{X} dt \cdot (\alpha M + \beta) = dt \cdot \beta + dt \cdot \left(\alpha M \frac{X - M}{X} \right) \\ &\approx dt \cdot \beta + dt \cdot (\alpha (X - M)) \end{aligned} \quad (5.34)$$

in case the population is not too overcrowded (i.e. M/X is close to 1).

Summarising, we establish the following model:

Model 2.4: Microscopic Levins Model with Immigration

The microscopic model (agent-based model) is defined by the following processes.

Initial Setup:

- Initially, generate a total of X_0 agents and for each of them, if possible, randomly assign a spot on the discrete line $\{1, \ldots, M\}$. If more than M agents are generated hereby, all agents which cannot be assigned a spot on the line are assigned the "overshoot" state. As long as there are agents with overshoot state in the model they are immediately assigned a regular spot on the line as soon as one becomes available during simulation.

The model is updated in discrete steps of length dt. All agent updates are done simultaneously.

Dynamics:

- Each step addresses each agent once. For each agent, pick a random spot on the discrete line $\{1, \ldots, M\}$. If the spot is not yet taken, the agent may give birth to a new agent with probability $dt \cdot M\alpha$. The "newborn agent" is moreover located at this spot.
- Moreover, the addressed agent dies with probability $dt \cdot \beta$ if it is a regular inhabitant and with probability $dt \cdot (\beta + \alpha M)$ if it is in the overshoot state. Hereby it is removed and its spot is freed.
- At the end of each time-step one additional agent may immigrate with probability $dt \cdot \gamma$. If no free spot is available for this agent, it is assigned the overshoot state.

The model processes are sketched in Figure 5.11.

Though the microscopic model definition seems strange, it can be interpreted from the modelling perspective quite nicely. The concept, that overpopulation has a negative influence on the survival of those who do not have access to all necessary resources is quite rational. This process is

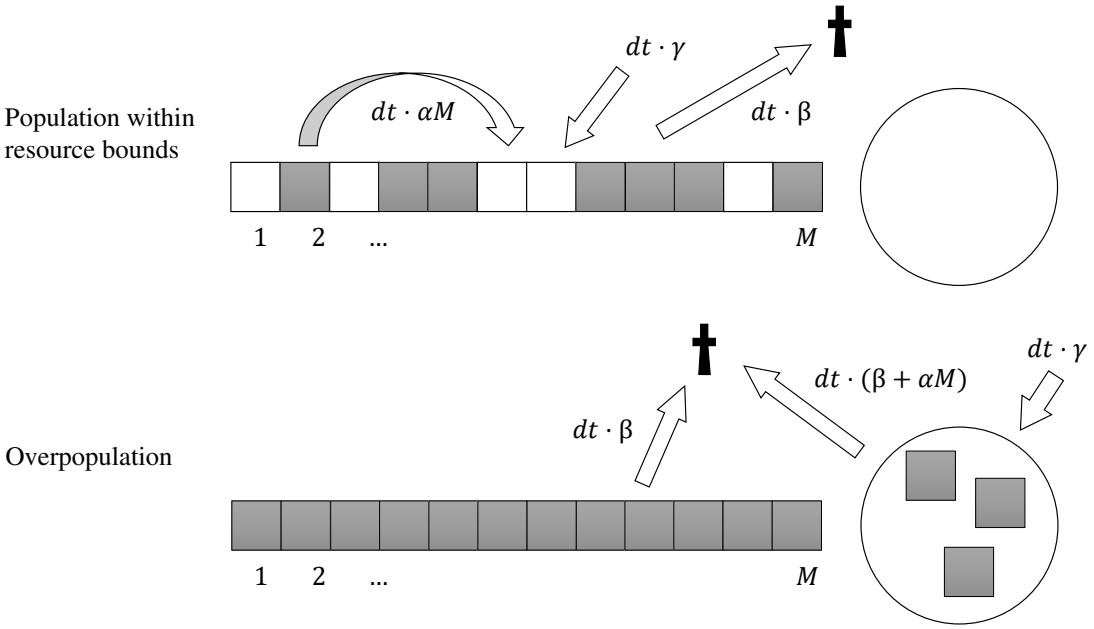


Figure 5.11: Sketch of the established microscopic Levins Model. While the population is within its resource-bounds (upper sketch), birth processes are possible. If the population is too large, the overshoot agents have an increased death probability.

present, but quite difficult to notice when only taking the macroscopic model into account. It is yet questionable if the death probability of the overcrowded population should have something to do with the birth probability of the population in bounds. In order to make the model more realistic, this point definitely leaves space for improvement.

Figure 5.12 shows a comparison of results of a single microscopic model run and the solutions of 5.27 for different parameters. The asymptotic equivalence between the two model results is obvious.

The stated model is one case study for finding a microscopic analogue for a macroscopic, equation-based model. One can, in general, distinguish two key issues, that motivate this inverse mean-field process.

The first one is the need for a microscopic representation for the sake of the model as a representation of the real system. I.e. for specific reasons, a microscopic representation is better suited for depicting the real system that is depicted by the macroscopic model. This issue often occurs in the course of a model extension or modification for which the inclusion of microscopic parameters or processes is necessary or, at least, beneficial. For example, if one wants to model heterogeneity within the population, the stated microscopic Levins model is definitely more flexible than the macroscopic one. Also, the high picturesqueness of microscopic models might be a reason why one wants to switch from a macroscopic to a microscopic approach. Hereby

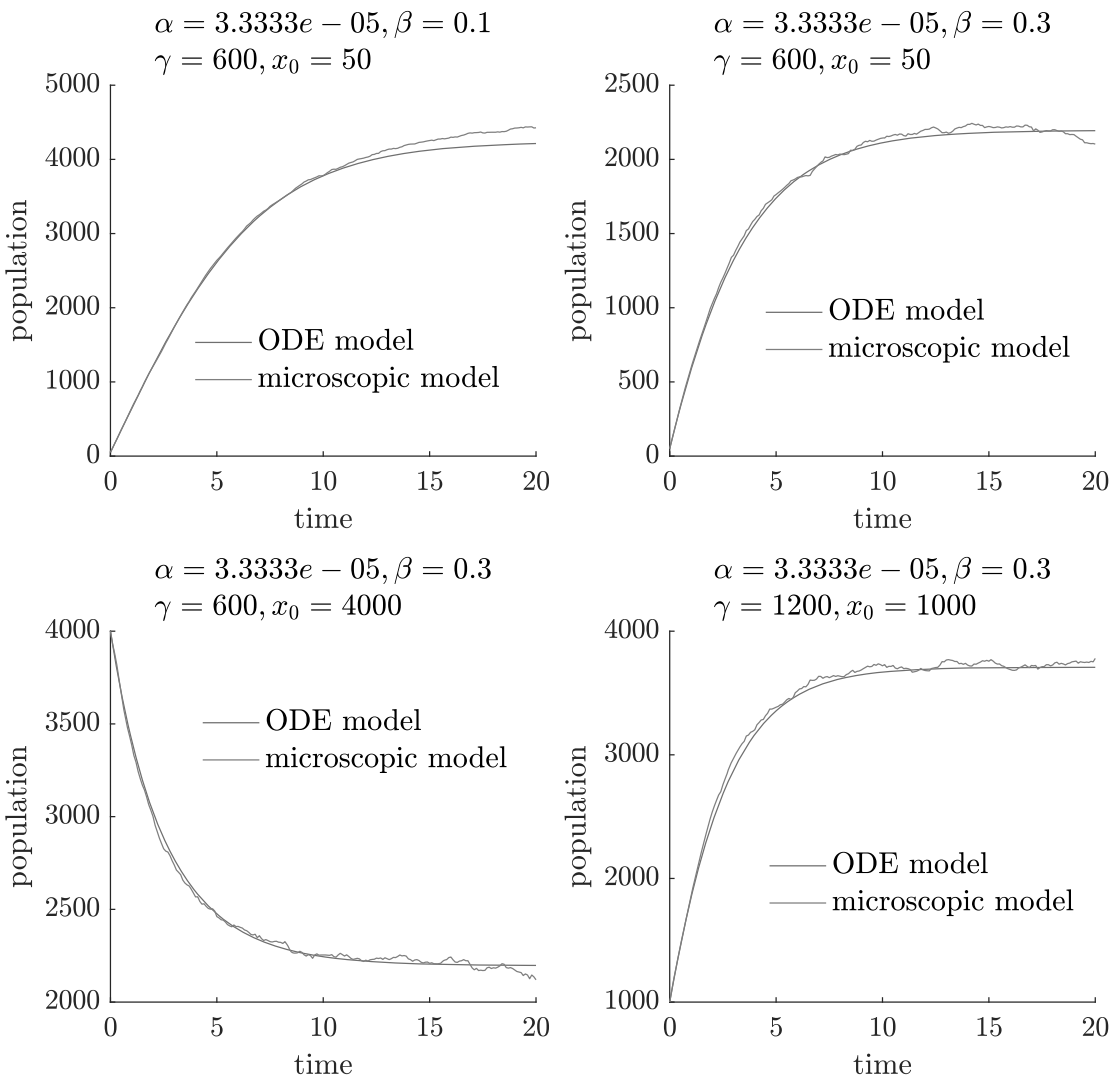


Figure 5.12: Levins ODE Model with immigration (5.27) in comparison with the defined agent-based model, Model 2.4, using different parameter settings. The upper-left picture shows that the simplification done in (5.34) disturbs the similarity of the two approaches if the population is too much overcrowded. Here, the microscopic model result drift off slightly.

probably a higher level of credibility can be achieved.

The second reason, why a microscopic representation might be superior to a macroscopic one relates to the technical issues with the simulation. Microscopic approaches are executed straight forwardly with simple loops and if-cases. Though they usually require comparably long computation times, their implementation is not related to any numerical difficulties which makes them a reasonable alternative to numerically challenging ODE or, especially, PDE models. The most famous examples are found in Physics. We like to mention the HPP [Hardy et al., 1976], FHP [Frisch et al., 1986] or Lattice-Boltzmann method [Chen and Doolen, 1998] which are proven to be microscopic alternatives to the famous Navier-Stokes equations for fluid flow (probably first published in 1845 by Stokes [Stokes, 1845]). While the simulation of the related PDE system is numerically challenging if the observed geometries, i.e. the shapes of obstacles and boundaries for the fluid, are complex, they do not pose any problems for the microscopic (particle-based) approaches.

5.3 Mean-Field Analysis for Lattice Models

As seen in the SIR model in Section 5.2.1 the key problem of performing a mean-field analysis (directly, by applying a mean-field theorem, or passively, by applying a step-by-step instruction) lies within identifying correct transition-rates or transition-probabilities as we usually have to find parts of the model that can be neglected.

In the mean-field analysis process of the SIR model presented in the last chapter, we were forced to neglect the agents' position on the grid to find a transition probability for the switch from the susceptible state to the infectious state. Hereby we assumed that the states among the agents are uniformly distributed. The last part of the case study (Figure 5.9) clearly underlines that this is not the case if there is no movement among the agents.

This section is dedicated to the analysis of this problem on more general *lattice-models*. With this term we denote any kind of microscopic model wherein a lattice is used as a spatial dimension. Essentially, cellular automata (Section 2.1.2) and agent-based models, wherein each agent is located on one of a finite number of grid-points, belong to this type.

5.3.1 Neighbourhood Concepts in Lattice Models

Though it was mentioned in Section 2.1.2 that a lattice or grid can be anything that can be indexed with whole numbers, we want to focus on finite, two-dimensional rectangular grids in this section – i.e.

$$Ix : \{1,\dots,m\} \times \{1,\dots,n\} \to C : \ (i,j) \mapsto Ix(i,j) = c_{i,j}.$$

Thinking of C as a spatial dimension, clearly some kind of distance metric would be useful. Hereby given two sites $c_{i,j}$ and $c_{k,l}$, intuitively three different metrics make sense:

$$d_1(c_{i,j}, c_{k,l}) := \|c_{i,j} - c_{k,l}\|_1 := |i-k| + |j-l|, \tag{5.35}$$

$$d_2(c_{i,j}, c_{k,l}) := \|c_{i,j} - c_{k,l}\|_2 := \sqrt{(i-k)^2 + (j-l)^2}, \text{and} \tag{5.36}$$

$$d_\infty(c_{i,j}, c_{k,l}) := \|c_{i,j} - c_{k,l}\|_\infty := \max(|i-k|, |j-l|). \tag{5.37}$$

These three norm-induced distance metrics are slightly more interesting on a lattice than on the classic $\mathbb{R}^2$ plane. This becomes clear when investigating the ε-spheres

$$U_\varepsilon(c_{i,j}) := \{c_{k,l} : d_x(c_{i,j}, c_{k,l}) \leq \varepsilon\} \tag{5.38}$$

as shown in Figure 5.13. Here ε-spheres with radius $\varepsilon = 1$ and $\varepsilon = 4$ are shown for all three mentioned distance metrics. For radius 1 the metrics only contain sites directly attached to the site in the middle. Thus, they are commonly denoted as (classic) **neighbourhood**. The first two of them result in the same neighbourhood, which is named Von Neumann neighbourhood after computer engineering pioneer John Von Neumann. He first used this metric in his famous 28-state cellular automaton (slightly more about this was given in Section 2.1.2). The third neighbourhood induced by the ∞-norm (which can also be called Chebyshev-distance) is named Moore neighbourhood after Edward F. Moore. He was, like Von Neumann, one of the founding fathers of cellular automata theory.

Circles with radii ≥ 2 contain sites that are not directly attached to the centre-site. They are often called indirect neighbours or, in case of the $d_1(\cdot)$ and $d_\infty(\cdot)$ metric, **neighbours of higher order**. Their neighbourhood order is usually defined slightly different: it results from finding chains of sites, so that the regarded site in the neighbourhood is in the neighbourhood of another site which is in the neighbourhood of another site, ..., which is in the neighbourhood of the centre-site. The minimal length of all possible chains is called order of the site's neighbourhood. The sites in the outermost rows or columns of $\{c_{k,l} : d_{1 \text{ or } \infty}(c_{i,j}, c_{k,l}) \leq k\}$, $k \in \mathbb{N}$ are neighbours of order k of $c_{i,j}$ with respect to the order Von Neumann/ Moore neighbourhood. Hence, for any $k \in \mathbb{N}$

$$\{c_{k,l} : d_{1(\text{or } \infty)}(c_{i,j}, c_{k,l}) \leq k\} \tag{5.39}$$

is called the **k$^{\text{th}}$ order Von Neumann/ Moore neighbourhood** of $c_{i,j}$.

Equipped with distance-metrics, it is a commonly used strategy in lattice models that the range of influence of a site or of an agent that is located at a site is limited by an ε-sphere – e.g. a neighbourhood or a higher order neighbourhood. Clearly, this concept may lead to local effects which cannot be ignored in a mean-field analysis process, as done by the "homogeneity assumption" in the microscopic SIR model. Is it even possible to perform mean-field approximations for these cases? We analyse this problem by a case study.

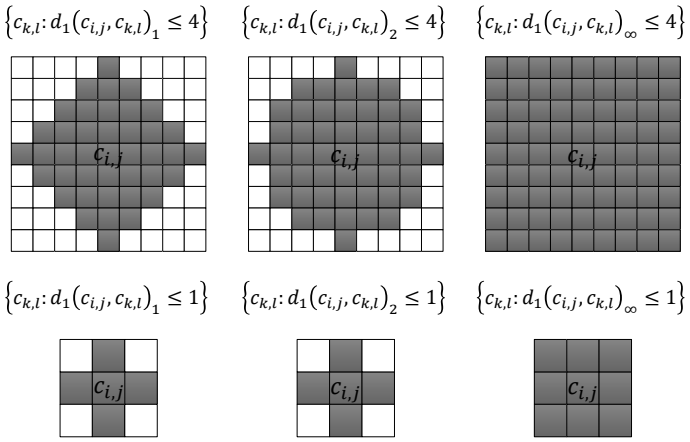


Figure 5.13: Circles with specific radius for all three specified lattice-metrics, with $c_{i,j}$ being the centred site. For radius 1, the $d_1(\cdot)$ metric and the $d_2(\cdot)$ metric lead to the Von Neumann neighbourhood, the last one $d_\infty(\cdot)$ results in the Moore neighbourhood.

5.3.2 Neighbourhood Case Study: Logistic Growth Lattice Model

In Section 4.6.5 we investigated a modified version of the so-called Levins model, a simplified macroscopic population model that includes a logistic growth of the population, linear death rates and an additional immigration term. In this section we will not care about immigration and death, but focus on the logistic growth. Note, that the lattice structure has no influence on death and immigration.

On the contrary to Section 4.6.5, a microscopic model will be the basis for the case study.

Model 3.1: Microscopic Growth Model

We define the following cellular automaton model, furthermore denoted as microscopic growth model, via its initial setup and the involved dynamics.

Initial Setup:

- First, a rectangular $N_1 \times N_2$ ($N_1 N_2 =: N$) grid C of sites $c_{i,j}, i \in \{1, \ldots, N_1\}, j \in \{1, \ldots, N_2\}$ is defined. A total number of $N \geq X_0$ randomly chosen sites are assigned state 1 (*inhabited*), to all others, state 0 (*empty*) is assigned.
- We pick a time-step size dt so that $\lambda \cdot dt$, whereas λ donates a birth parameter, is $\ll 1$. If $\lambda \ll 1$ we usually choose $dt = 1$.

The model-time is enhanced with discrete time-steps of length dt. In each step the following dynamics are executed.

Dynamics:

- Each site is addressed once.
- In case the addressed site $c_{i,j}$ is empty (0), a second site is picked randomly (uniformly) from a set site-specific set of sites $S(c_{i,j})$. In case the picked site is inhabited (1), the addressed site becomes inhabited with probability $\lambda \cdot dt$. Hereby the inhabited site "gave birth to an offspring".

Aim of the following analysis is to investigate different choices of $S(c_{i,j})$. We will start with the rather unnatural idea that an individual may reproduce anywhere on the total grid $S(c_{i,j}) := C$, but we aim to analyse more realistic measures, e.g.

$$S(c_{i,j}) := \{c_{k,l} : d_{1(\text{or }\infty)}(c_{i,j}, c_{k,l}) \leq k\}$$

for some k or even more sophisticated "picking concepts".

- **Scenario base**: S assigns a random site in the whole site-space:

$$S(c_{i,j}) = C,$$

- **Scenario a**: S assigns all sites within Euclidean distance: for $k \geq 2$:

$$S(c_{i,j}) := \{c_{k,l} : d_2(c_{i,j}, c_{k,l}) \leq r\},$$

- **Scenario b**: S assigns all sites in Moore neighbourhood:

$$S(c_{i,j}) := \{c_{k,l} : d_\infty(c_{i,j}, c_{k,l}) \leq 1\},$$

- **Scenario c**: S assigns all sites in Von Neumann neighbourhood:

$$S(c_{i,j}) := \{c_{k,l} : d_1(c_{i,j}, c_{k,l}) \leq 1\},$$

- **Scenario d**: S only assigns the site directly to the right:

$$S(c_{i,j}) := \{c_{i+1,j}\}$$

, and

- **Scenario e**: S assigns both, the site directly to the right and to the left:

$$S(c_{i,j}) := \{c_{i+1,j}, c_{i-1,j}\}.$$

Hereby all metrics are extended toroidally beyond the borders of the grid – that means, e.g. $d_1(c_{N_1,j}, c_{1,j}) = 1$. Analogously, for Scenario d and e, the rules $c_{1,j+1} \in S(c_{N_1,j})$ and $c_{1,1} \in S(c_{N_1,N_2})$ are defined. Thus, for those two scenarios, the cellular automaton can be imagined as a one-dimensional line connected at its ends – i.e. a ring. We will also call Scenario d and e as **one- and two- sided 1-dimensional neighbourhood**. For Scenario a we usually refer to parameter r as **neighbourhood-radius** for its similarity with circles.

Using the step-by-step instruction in Section 4.6.4 it is a simple task to prove that the classic Logistic equation

Model 3.2: Logistic Growth ODE

$$\frac{d}{dt}\varphi(t) = \lambda\varphi\frac{N-\varphi}{N}, \quad \varphi(0) = X_0 \tag{5.40}$$

poses for a good mean-field approximation for the total number of inhabited sites

$$X_1(t) := \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} \mathbb{1}_1 c_{i,j}(t)$$

in the base scenario. We need to consider the probability that a site with state 0 changes its status to 1 in a time-step: As a random contact partner is drawn from the total site-space, it is an inhabited one with probability $\frac{X_1(t)}{N}$. Moreover, the contact results in an offspring with probability $dt \cdot \lambda$. Hence

$$\frac{P(I_i(t+dt)=1|I_i(t)=0)}{dt} = \frac{1}{dt}\frac{X(t)}{N}dt\cdot\lambda = \frac{\lambda X(t)}{N}.$$

Let indexing by 1 and 0 denote the total number of inhabited and empty sites, the corresponding mean-field equation is given as follows

$$\frac{d}{dt}\begin{pmatrix}\varphi_1(t)\\ \varphi_0(t)\end{pmatrix} = \begin{pmatrix}\varphi_0(t)\frac{P(I_i(t+dt)=1|I_i(t)=0)}{dt}\\ -\varphi_0(t)\frac{P(I_i(t+dt)=1|I_i(t)=0)}{dt}\end{pmatrix} = \begin{pmatrix}\lambda\varphi_0(t)\frac{\varphi_1(t)}{N}\\ -\lambda\varphi_0(t)\frac{\varphi_1(t)}{N}\end{pmatrix}.$$

As $\varphi_0 = N - \varphi_1$ for balance reasons, the second equation is redundant and the first one becomes equivalent to the stated logistic model.

The derived mean-field model is well known and does not surprise anyone who has read Section 5.2.3. Yet things tend to become more delicate when more local picking scenarios are used. Taking a closer more analytic look at the defined model scenarios, the experienced modeller may get some first ideas about the expected model outcome without even simulating the model:

For all model scenarios the population size is almost surely growing until all sites are occupied. Nevertheless, it is a question of the used neighbourhood, how fast it grows. As a smaller neighbourhood size reduces the connectivity of the sites in the automaton, it is logical to suggest, that the smaller the number of regarded sites, the slower is the growth of the population. This suggestion can be seen to hold when comparing results of the model for the different scenarios. These can be seen in Figure 5.14. Hereby the sample mean of 100 simulation-runs per Scenario are plotted against each other. For Scenario a, two different neighbourhood radii $r = 7$ and $r = 3$ were chosen for comparison. Obviously the growth of the population is slower, the smaller the number of neighboured sites. For neighbourhood radius $r = 7$, i.e. 148 sites in the ε-sphere, the results are almost equivalent to the results of the base scenario. The results of the logistic ODE (5.40) can be seen to be a good fit to these two scenarios. Surprisingly the results from Scenario e can hardly be distinguished from Scenario d. This result is unexpected and unintuitive: the population in the model with two neighbours spreads as fast as the population with only one neighbour. We try to find the reason for the latter phenomenon using mean-field analysis.

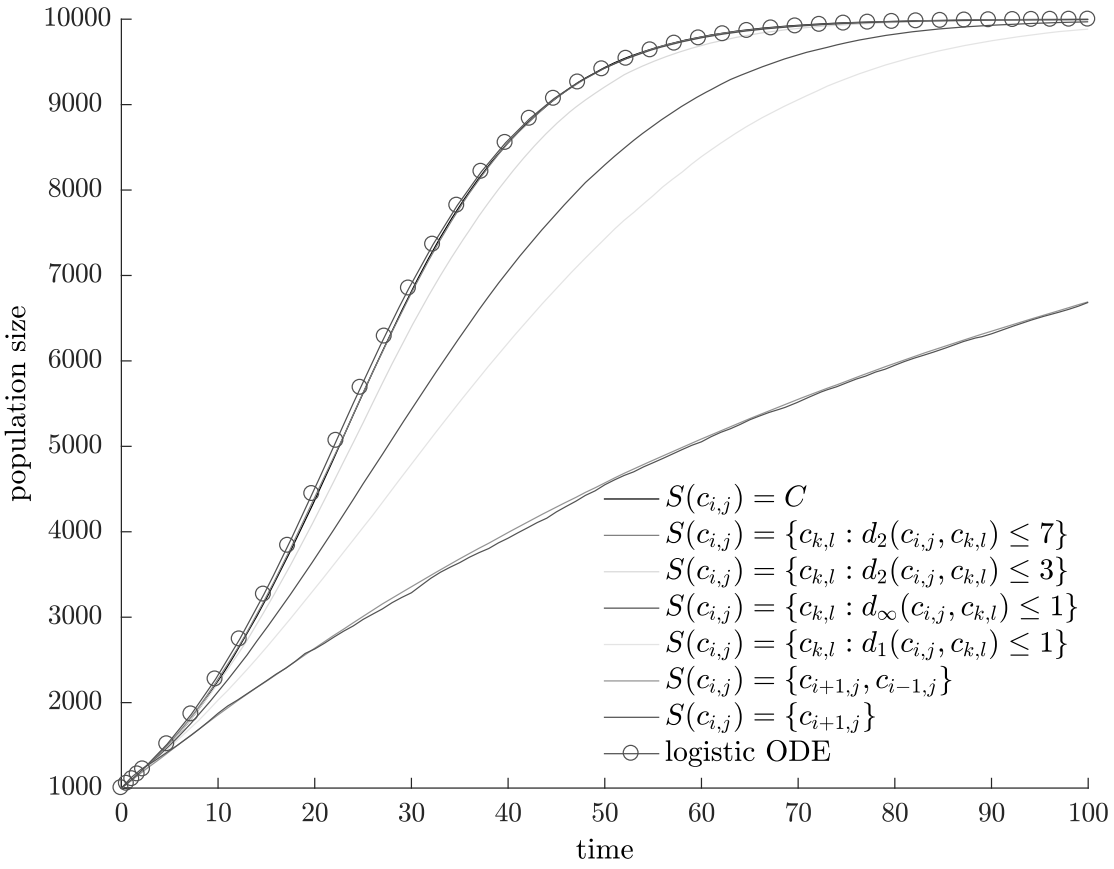


Figure 5.14: Sample result of the microscopic growth model 3.1 for different neighbourhood concepts. A grid with $N_1 = N_2 = 100$ sites was used. Hence, the fluctuations of these curves are expected to have asymptotic size of $\mathcal{O}(\sqrt{10000}) = \mathcal{O}(100)$ and can be neglected.

5.3.3 1D-Neighbourhood Case Study: Logistic Growth Lattice Model

We will **start with a mean-field analysis of Scenario d** and regard Scenario e afterwards.

As mentioned before, it is clear, that the position of the site inside the grid can only be neglected for the case $S(c_{i,j}) = C$ or if the number of neighbours is large enough to pose for a snipped of the grid that is somehow representative with respect to distribution of 0-s and 1-s. In that case we are able to use the concept of Laplace space in the mean-field analysis and receive the standard Logistic Growth ODE. Especially in the case of the one- and two- sided 1d-neighbourhoods, this is not possible as the effects are simply too local. Yet, based on the techniques stated in [Fibich and Gibori, 2010] we found a method to determine a correct mean-field curve for it.

First, we write

$$\hat{c}_i := c_{j,k}, \quad i = j + k(N_1 - 1)$$

to reorder the sites so that they are useful for 1-dimensional analysis. Hereby the notation becomes considerably easier. Note, that the site to the right of $\hat{c}_N$ is $\hat{c}_1$.

Moreover, we already noticed, that the state-space $\Gamma = \{0, 1\}$ is not sufficient to depict the behaviour of the automaton. Hence, we decided to enlarge not only the state-space, but change the view on the sub-models of the model:

We do not only consider all sites in C but regard the set of all possible aligned tuples

$$C' := \{[\hat{c}_i, \ldots, \hat{c}_{i+j}], j \in \{0, \ldots, N\}\}$$

as the microscopic parts of the models. Clearly, the minimal state-space Γ' that covers all states of all tuples can be characterised as the set of all possible $0 - 1$ combinations with any length smaller than N:

$$\Gamma' := \{[x_1, \ldots, x_j] : \forall k : x_k \in \{0, 1\}, j \in \{0, \ldots, N\}\}.$$

As usual, we write (exemplary for the tuple $[1011]$)

$$X_{[1011]} := \sum_{[\hat{c}_i, \ldots, \hat{c}_{i+j}] \in C'} \mathbb{1}_{[1011]}([\hat{c}_i(t), \ldots, \hat{c}_{i+j}(t)])$$

for the aggregated numbers of the model with respect to the newly defined state-space. Hence, $X_{[1011]}(t)$ counts the total number of (aligned) quadruples of sites, that have the state-combination $[1, 0, 1, 1]$ at time t.

Clearly, the huge state-space Γ' is much too large to be analysed fully. Yet, regarding the target value, namely the total number of sites in state 1, this is not necessary. For finding a closed formulation of a mean-field model for it (only) the following aggregated numbers need to be analysed:

- $X_{[1]}(t) := \sum_{i=1}^{N} \mathbb{1}_1(\hat{c}_i(t))$.

  Hence, $X_{[1]}(t)$ denotes the total number of states 1 in the automaton. It is important to understand, that this **is** our target value.

- $X_{[0]}(t) := \sum_{i=1}^{N} \mathbb{1}_0(\hat{c}_i(t)) \quad (= N - [1])$.

  Hence, $X_{[0]}(t)$ denotes the total number of states 0 in the automaton. As each site can either be 0 or 1 this fraction is complementary to $X_{[1]}$.

- $X_{[0k]}(t) := X_{[\underbrace{0 \ldots 0}_{k\times}]}(t) := \sum_{[\hat{c}_i, \ldots, \hat{c}_{i+k-1}] \in C'} \mathbb{1}_{[0, \ldots, 0]}([\hat{c}_i(t), \ldots, \hat{c}_{i+k-1}(t)])$.

  Hence, $X_{[0k]}$ denotes the total number of sequences of states 0 with length k in the lined up automaton. It counts the number of k consecutive 0-s.

- $X_{[0k1]}(t) := X_{[\underbrace{0 \ldots 0}_{k\times} 1]}(t) := \sum_{[\hat{c}_i, \ldots, \hat{c}_{i+k}] \in C'} \mathbb{1}_{[0, \ldots, 0, 1]}([\hat{c}_i(t), \ldots, \hat{c}_{i+k}(t)])$.

  Hence, $X_{[0k1]}$ counts the fraction of k consecutive 0-s followed by a 1 in the automaton.

For space reasons we furthermore drop the dependence of t in our notations. We take a look at the resulting equations in the Discrete Space Mean-Field Theorem (MFT) with Static Population and get the following two mean-field differential equations for the expected values of $X_{[1]}$ and $X_{[0]}$:

$$\varphi'_{[1]} = \varphi_{[0]} \frac{P(0 \to 1)}{dt}, \quad \varphi'_{[0]} = -\varphi_{[0]} \frac{P(0 \to 1)}{dt}.$$

Note, that there is no transition that leads from 1 back to 0. Clearly, the probability can be split into two parts: The probability that the next site is in state 1 and the probability for a state-switch (multiplied by dt as we need to approximate a rate and not a probability):

$$P(0 \to 1) = P(\text{the site is followed by a site with state } 1)\lambda \cdot dt.$$

Thus, the target site with state 0 needs to be part of a tuple $[01]$. By the laws of conditional probabilities, the likelihood that a randomly picked site with state 0 is the first part of one of these tuples is given by the total amount of tuples $[01]$, i.e. $X_{[01]}$ divided by the total amount of sites with state 0, i.e. $X_{[0]}$. It follows that

$$\frac{P(0 \to 1)}{dt} = \frac{X_{[01]}}{X_{[0]}} \frac{\lambda \cdot dt}{dt} = \frac{X_{[01]}}{X_{[0]}} \lambda,$$

which leads to

$$\varphi'_{[1]} = \lambda \varphi_{[0]} \frac{\varphi_{[01]}}{\varphi_{[0]}} = \lambda \varphi_{[01]}, \quad \varphi'_{[0]} = -\lambda \varphi_{[0]} \frac{\varphi_{[01]}}{\varphi_{[0]}} = -\lambda \varphi_{[01]}. \tag{5.41}$$

In order to create a relationship between $[01]$ and $[0k]$ we calculate

$$X_{[0]} = X_{[00]} + X_{[01]}.$$

This is true as any 0 that contributed to $X_{[0]}$ is either followed by another 0 which contributes to $X_{[00]}$, or by a 1, contributing to $X_{[01]}$. Consequently, $\varphi_{[0]} = \varphi_{[00]} + \varphi_{[01]}$ needs to be fulfilled for the expected values in the mean-field model as well. We can generalise this idea by

$$X_{[0k]} = X_{[0k1]} + X_{[0(k+1)]} \Rightarrow \varphi_{[0k]} = \varphi_{[0k1]} + \varphi_{[0(k+1)]}. \tag{5.42}$$

We furthermore use the Discrete Space MFT with Static Population to derive differential equations for the states $X_{[0k]}$. As before, the probability that a tuple $[0k]$ transforms into $[0(k-1)1]$ in a time-step is given by the product of $\lambda \cdot dt$ and the probability of the $[0k]$ tuple to be part of a $[0k1]$ tuple. The latter is given by the total number of $[0k1]$ tuples divided by the total number of $[0k]$ tuples. Hence,

$$\varphi'_{[0k]} = -\varphi_{[0k]} \frac{P([0k] \to [0(k-1)1])}{dt} = -\varphi_{[0k]} \lambda \frac{\varphi_{[0k1]}}{\varphi_{[0k]}} = -\lambda \varphi_{[0k1]}. \tag{5.43}$$

As the initial setting is uniformly distributed we get $\varphi_{[0k]}(0) = \frac{(N-X_0)^k}{N^{k-1}}$. Combining equations (5.42) and (5.43) we finally derive the following (basically finite) set of differential equations:

$$\left\{ \varphi'_{[0k]} = -\lambda \varphi_{[0k]} + \lambda \varphi_{[0(k+1)]}, \quad \varphi_{[0k]}(0) = \frac{(N-X_0)^k}{N^{k-1}} \right\}_{k=1}^{N-X_0} \tag{5.44}$$

It is obvious that $\forall k > N - X_0 : \ \varphi_{[0(k)]}(t) = 0$, as already the initial number of $[0]$-s, namely $N - X_0$, is smaller than the zeros in the regarded state-combination.

As N is usually a large number it is not useful to solve this huge system of coupled equations directly. It is surprisingly more useful to approximate the finite system with an infinite one:

$$\left\{ \varphi'_{[0k]} = -\lambda\varphi_{[0k]} + \lambda\varphi_{[0(k+1)]}, \quad \varphi_{[0k]}(0) = \frac{(N - X_0)^k}{N^{k-1}} \right\}_{k=1}^{\infty} \tag{5.45}$$

The error clearly converges to zero, when N goes to infinity. By induction we claim a solution for the infinite system by

$$\varphi_{[0k]}(t) := \lambda \frac{(N - X_0)^{k-1}}{N^{k-1}} \varphi_{[0]}(t), \tag{5.46}$$

for $k > 1$, if $\varphi_{[0]}$ solves

$$\varphi'_{[<0]} = -\lambda\varphi_{[0]} + \lambda\varphi_{[00]} = -\lambda\varphi_{[0]} + \frac{\lambda(N - X_0)}{N}\varphi_{[0]} = -\lambda\frac{X_0}{N}\varphi_{[0]}, \quad \varphi_{[0]}(0) = (N - X_0) \tag{5.47}$$

So, the set of functions

$$\varphi_{[0k]}(t) = e^{-\frac{X_0\lambda t}{N}} \frac{(N - X_0)^k}{N^{k-1}}, \ k \in \mathbb{N} \setminus \{0\}$$

solves (5.45).

This is one of the very rare applications for infinite systems of ODEs and theory about the uniqueness of solutions is not trivial. Hence, we decided to exclude possible alternative solutions. We think of a second solution $\hat{\varphi}_{[0k]}$ and calculate:

$$(\varphi_{[0k]} - \hat{\varphi}_{[0k]})' =: a'_k := -\lambda a_k + \lambda a_{k+1}, \quad a_k(0) = 0.$$

A substitution $b_k := e^{\lambda t} a_k$ leads to

$$b'_k = \lambda e^{\lambda t} a_k + e^{\lambda t} a'_k = \lambda e^{\lambda t} a_k - \lambda e^{\lambda t} a_k + \lambda e^{\lambda t} a_{k+1} = \lambda b_{k+1}$$

and furthermore

$$b_k^{(n)} = \lambda^{n-k} b_{n+1}.$$

So for any positive k and n we receive $b_k^{(n)}(0) = \lambda^{n-k} b_{n+1}(0) = 0$ which leads to the Taylor series decomposition:

$$b_k(t) = b_k(0) + t b_k^{(1)}(0) + \frac{t^2}{2} t b_k^{(2)}(0) + \cdots = 0.$$

We receive $b_k \equiv 0$, $e^{\lambda t} a_k \equiv 0$, $a_k \equiv 0$ and hence $\hat{\varphi}_{[0k]} \equiv \varphi_{[0k]}$ for all k. Thus, the solution must be unique.

As

$$\varphi'_{[0]} = -\frac{\lambda X_0}{N}\varphi_{[0]} \Rightarrow \varphi'_{[1]} = \frac{\lambda X_0}{N}(N - \varphi_{[1]})$$

we found that the target aggregated number $X_{[1]}$ is solved by a linear ordinary differential equation.

We will now regard Scenario d. To finally compare the one-sided with the two-sided neighbourhood to find the reason, why the model curves are almost identical, we perform a similar calculation for the two-sided case. Using the same notation as before we get

$$\varphi'_{[0]} = -\varphi_{[0]} \frac{(\lambda \frac{\lambda}{2} \varphi_{[100]} + \varphi_{[101]} + \frac{\lambda}{2} \varphi_{[001]})}{\varphi_{[0]}}.$$

As there are always the same amount of $[001]$ as $[100]$ combinations in the model, we get

$$\varphi'_{[0]} = -\lambda(\varphi_{[101]} + \varphi_{[001]}).$$

As $X_{[101]} + X_{[001]} = X_{[01]}$ and therefore $\varphi_{[101]} + \varphi_{[001]} = \varphi_{[01]}$,

$$\varphi'_{[0]} = -\lambda \varphi_{[01]} \tag{5.48}$$

results. Analogously

$$\varphi'_{[0k]} = -\varphi_{[0]} \frac{(\lambda \varphi_{[10k1]} + \frac{\lambda}{2} \varphi_{[10k0]} + \frac{\lambda}{2} \varphi_{[00k1]})}{\varphi_{[0]}} = -\lambda(\varphi_{[10k1]} + \varphi_{[00k1]}) = -\lambda \varphi_{[0k1]} \tag{5.49}$$

follows. Altogether (5.48) and (5.49) result in the same system of differential equations as in the one-sided case via (5.41) and (5.43). As also the initial conditions are equivalent, both systems are solved by the same solution curves.

Finally, mean-field analysis proves, that **Scenario d and e behave alike on the mean-field level**. Hence,

Model 3.3: Linear Growth ODE

$$\frac{d}{dt}\varphi(t) = \lambda X_0 \frac{N - \varphi}{N}, \quad \varphi(0) = X_0 \tag{5.50}$$

is a valid mean-field approximation for Scenarios d **and** e. This can be seen to hold in Figure 5.15, which is in principle a copy of Figure 5.14 with the additional solution curve of the Linear Growth ODE model. Note, that the linear growth model 3.3 must not be confused with an exponential growth model, as it is a bounded growth to a steady-state. Figure 5.16 shows a direct comparison of the higher order quantities $[0k]$ in the one-sided 1-d model: left, counted quantities in the microscopic model are shown, right, solutions of the ODE system (5.45).

5.3.4 Pair Approximation Case Study: Logistic Growth Lattice Model

Results gained from the previously performed mean-field approximations showed two limiting curves for all applied neighbourhood scenarios. The first one, the logistic equation (5.40), describes the behaviour of the fully random contact model. It is a pretty simple task to show that for any more-restricted neighbourhood the speed of the growth process is slowed down. We will

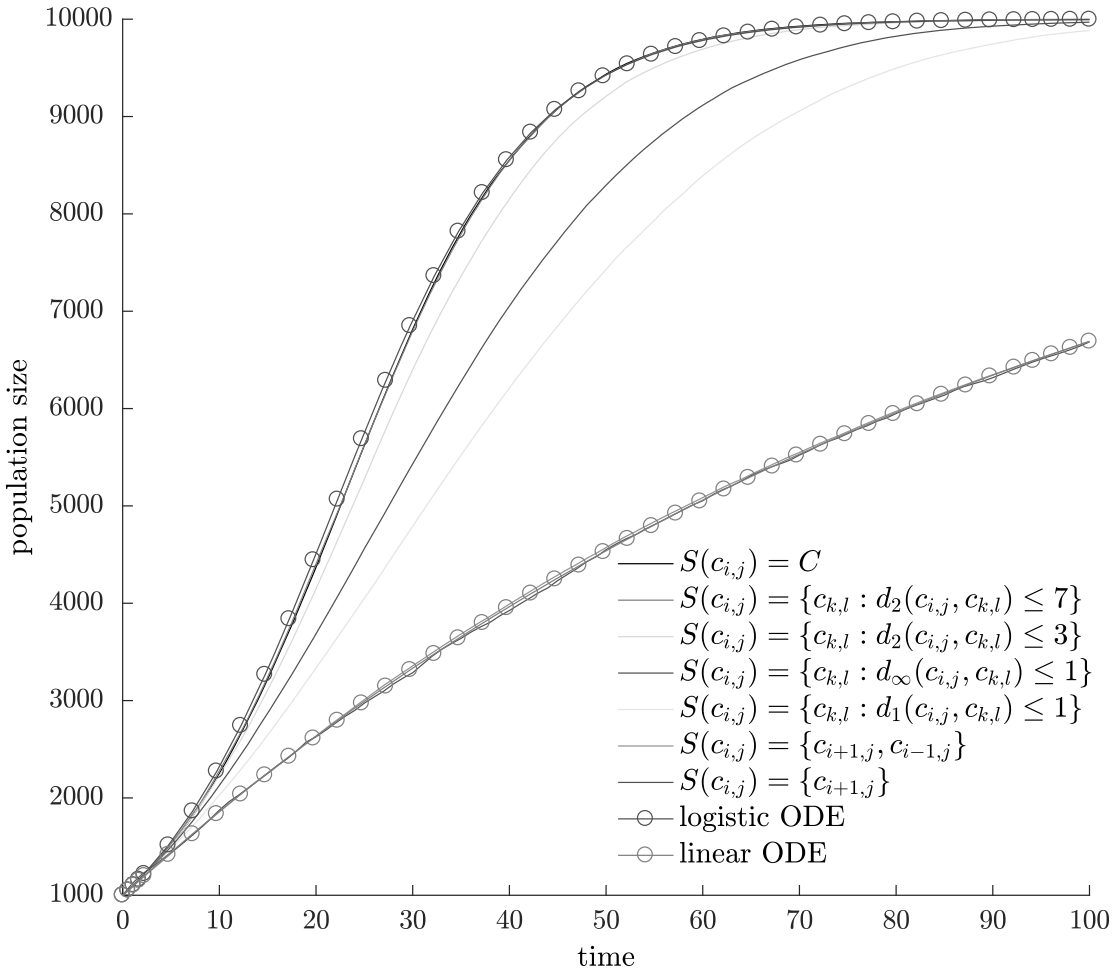


Figure 5.15: Sample plot as Figure 5.14 included the .

not go into details, but refer to [Fibich and Gibori, 2010] for a proof. Hence, the logistic equation is somehow an upper bound for all mean-field curves of all other neighbourhoods. Analogously the linear ODE (5.50) gained from the 1-d neighbourhood poses for a lower bound for all other mean-field curves.

As a result candidates for a mean-field model for Scenarios a,b and c necessarily have to qualitatively lie somewhere in the middle between the linear and the non-linear model, i.e. in the middle between

$$\frac{d}{dt}\varphi(t) = \lambda X_0 \frac{N-\varphi}{N}, \quad \varphi(0) = X_0$$

and

$$\frac{d}{dt}\varphi(t) = \lambda \varphi \frac{N-\varphi}{N}, \quad \varphi(0) = X_0.$$

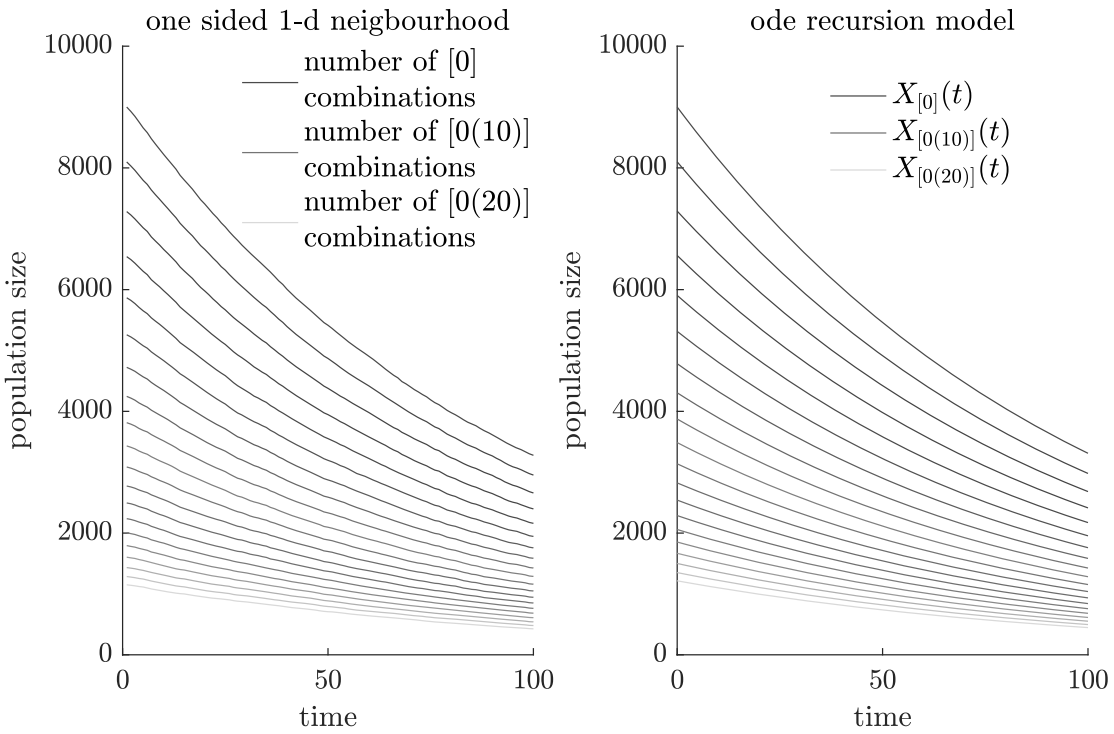


Figure 5.16: Comparison of the results of the one-sided 1-d neighbourhood growth model (Model 3.1, Scenario e) and the results of the derived ODE recursion (5.45). The counted quantities of $[0k]$ combinations for k from 1 to 20 in the microscopic model are shown to the left, the corresponding ODE solution curve is shown to the right.

The fact that a turning point in the solution curves of Scenarios b and c is well visible indicates, that any mean-field approximation of these two scenarios is either non-linear or, in case it is linear, has at least order 2.

On the one hand, like the 1-d neighbourhood concepts, it is obvious that it is not possible to perform a mean-field analysis by investigating the sites without considering their local environment. On the other hand, performing a similar concept to the 1-d neighbourhood, i.e. essentially enlarging the common state-space from $\Gamma = \{0, 1\}$ to some (large) subset of Γ^C leads to an un-treatable overhead. As a compromise, we introduce the technique of **pair approximation (PA)** applied to this specific example.

Pair Approximation or, to distinguish it from newer extensions, uncorrelated pair approximation (**UPA**), is a mean-field description technique that arose in the early 1990 for ecological models. It is commonly stated that the technique was first described in [Matsuda et al., 1992] wherein the method was applied to validly describe the mean-field behaviour of a microscopic predator-prey model on a lattice. The concept became quite successful and was applied to several different models. Even similar models as the automaton stated above have already been

analysed using pair approximation (e.g. [Ellner, 2001]).

The key idea can be imagined as follows: In principle, classical mean-field analysis as presented in Chapter 4 is applied to a set of sub-models each having a time-dependent state in a common state-space. In contrast PA does not apply the mean-field analysis on the obvious sub-model set

$$\{I_i : i \in \{1, \dots, N\}\} \tag{5.51}$$

with common state-space Γ, but on the sub-model set

$$\{(I_i, I_j) : I_i \sim I_j,\ i, j \in \{1 \dots, N\}\}, \tag{5.52}$$

with some relation $\sim$, and a common state-space

$$\Gamma' := \Gamma^2. \tag{5.53}$$

Classically, mentioned relation is given by a neighbourhood or neighbourhood-metric. For example, $I_i \sim I_j \Leftrightarrow d_1(I_i, I_j) \leq k$ for some k. By convention, we will refer to the first member of the tuple as the "focused" sub-model, while the second one is one of its possible contact partners.

Moreover, instead of calculating the transition probabilities

$$P(s_i \to s_j),$$

as done in standard mean-field analysis, PA requires calculation of the transition probabilities of the pairs

$$P((s_i, s_j) \to (s_k, s_l)).$$

To apply the pair approximation on Scenarios a,b and c we, first, define z as the total number of sites in any site's neighbourhood. I.e. $z = 8$ for scenario b (Moore neighbourhood), $z = 4$ for Scenario c (Von Neumann neighbourhood) and, exemplary, $z = 148$ for Scenario a with $r = 7$. Moreover, the following terms are fixed to simplify the formulas

$$c'_{i,j,k,l}(t) := (c_{i,j}(t), c_{k,l}(t)) \in \{0, 1\}^2, \quad \forall i, j, \text{ and } \forall k, l :\ c_{k,l} \in S(c_{i,j}) \tag{5.54}$$

$$X_{11}(t) := \sum_{i,j=1}^{N_1,N_2} \sum_{(k,l):\ c_{k,l} \in S(c_{i,j})} \mathbb{1}_{(1,1)} c'_{i,j,k,l}(t) \tag{5.55}$$

$$X_{10}(t) := \sum_{i,j=1}^{N_1,N_2} \sum_{(k,l):\ c_{k,l} \in S(c_{i,j})} \mathbb{1}_{(1,0)} c'_{i,j,k,l}(t) \tag{5.56}$$

$$X_{01}(t) := \sum_{i,j=1}^{N_1,N_2} \sum_{(k,l):\ c_{k,l} \in S(c_{i,j})} \mathbb{1}_{(0,1)} c'_{i,j,k,l}(t) \tag{5.57}$$

$$X_{00}(t) := \sum_{i,j=1}^{N_1,N_2} \sum_{(k,l):\ c_{k,l} \in S(c_{i,j})} \mathbb{1}_{(0,0)} c'_{i,j,k,l}(t) \tag{5.58}$$

Consequently, the four aggregated numbers $X_{m,n},\ m,n \in \{0,1\}$ denote the total number of pairs so that the first site, the focused one, currently has state m and the second one has state n at time t. As before, we will drop the dependency of t to save space. Note, that by $|S(c_{i,j})| = z$ for all i,j we hereby investigate a model with zN individual pairs.

Before going into details about how the defined aggregated numbers can be approximated, we discuss, how our target valued $X_1(t) = \sum_{i,j=1}^{N_1,N_2} \mathbb{1}_1(c_{ij}(t))$ can be derived from knowing the values of X_{11}, X_{10}, X_{01} and X_{00}. We get

$$\begin{aligned}
X_{11} + X_{10} &= \sum_{i,j=1}^{N_1,N_2} \sum_{(k,l):\ c_{k,l}\in S(c_{i,j})} \mathbb{1}_{(1,1)} c'_{i,j,k,l} + \sum_{i,j=1}^{N_1,N_2} \sum_{(k,l)\ :c_{k,l}\in S(c_{i,j})} \mathbb{1}_{(1,0)} c'_{i,j,k,l} \\
&= \sum_{i,j=1}^{N_1,N_2} \sum_{(k,l):\ c_{k,l}\in S(c_{i,j})} \mathbb{1}_1(c_{i,j})\mathbb{1}_1(c_{k,l}) + \sum_{i,j=1}^{N_1,N_2} \sum_{(k,l):\ c_{k,l}\in S(c_{i,j})} \mathbb{1}_1(c_{i,j})\mathbb{1}_0(c_{k,l}) \\
&= \sum_{i,j=1}^{N_1,N_2} \mathbb{1}_1(c_{i,j}) \sum_{(k,l):\ c_{k,l}\in S(c_{i,j})} (\mathbb{1}_1(c_{k,l}) + \mathbb{1}_0(c_{k,l})) \\
&= \sum_{i,j=1}^{N_1,N_2} \mathbb{1}_1(c_{i,j}) \sum_{(k,l):\ c_{k,l}\in S(c_{i,j})} 1 = \sum_{i,j=1}^{N_1,N_2} \mathbb{1}_1(c_{i,j}) z = z X_1.
\end{aligned}$$

The analogous result can be gained for X_0. Summarising,

$$X_1 = \frac{1}{z}(X_{11} + X_{10}), \quad X_0 = \frac{1}{z}(X_{01} + X_{00}). \tag{5.59}$$

Thus, the sum of the first two aggregated numbers results in X_1 scaled with the size z of the neighbourhood.

Moreover, as the used neighbourhood in Scenarios a, b and c are symmetric,

$$X_{10} = X_{01} \tag{5.60}$$

holds which may help to derive the mean-field equations. We do that by first identifying all possible transitions. Based on the model definition only the following four transitions are possible:

$$P((0,0) \to (0,1)), P((0,0) \to (1,0)), P((0,1) \to (1,1)), P((1,0) \to (1,1)).$$

Hence, a pair with state $(0,0)$ can only switch to $(0,1)$ or $(1,0)$ and a pair with state $(0,1)$ or $(1,0)$, respectively, can only switch to $(1,1)$. We derive the probabilities for these switches. Key to do that – **and most important feature of the PA** – is the fact, that the state of one of the neighbours of a site is already known, which has a positive or negative influence on the transition probability. As the pick among the neighbours is uniformly, the chance of picking the known neighbour is $1/z$ while there is the probability of $(z-1)/z$ to pick one of the other, unknown

neighbours. Hence, laws for conditional probabilities can be exploited.

$$\begin{aligned}
P((0,0)\to(1,0)) &= P\left(\text{the site switches to 1}\,\middle|\,\begin{array}{c}\text{the site has state 0,}\\ \text{a neighbour has state 0}\end{array}\right)\\
&= P(\text{switch})P\left(\begin{array}{c}\text{the picked neighbour}\\ \text{has state 1}\end{array}\,\middle|\,\begin{array}{c}\text{the site has state 0,}\\ \text{a neighbour has state 0}\end{array}\right)\\
&= P(\text{switch})\frac{z-1}{z}P(\text{the picked, unknown, neighbour has state 1}|\text{ the site has state 0})\\
&= \lambda\cdot dt\frac{z-1}{z}P(\text{the picked, unknown, neighbour has state 1}|\text{ the site has state 0}).
\end{aligned}$$

The second equality holds as randomly picking the known neighbour with state 0 does not allow the site to switch. Hence, only picking one of the $z-1$ known neighbours makes a switch possible. The remaining probability can be calculated considering the Laplacian formula: the probability to find a 1 as a neighbour of a 0 site is equivalent with the fraction of $(0,1)$ tuples (i.e. tuples with an empty site in the first coordinate and an inhabited neighbour site) divided by the total numbers of tuples with 0 on the first coordinate – i.e. $X_{01}+X_{00}$.

$$P((0,0)\to(1,0)) = \lambda\cdot dt\frac{z-1}{z}\frac{X_{01}}{X_{00}+X_{01}}. \tag{5.61}$$

This equality also assumes that the 01 and 00 combinations are somehow uniformly spread on the grid – we will see later, that this is not perfectly the case. Analogously,

$$\begin{aligned}
P((0,1)\to(1,1)) &= P\left(\text{the site switches to 1}\,\middle|\,\begin{array}{c}\text{the site has state 0,}\\ \text{a neighbour has state 1}\end{array}\right)\\
&= P(\text{switch})P\left(\begin{array}{c}\text{the picked neighbour}\\ \text{has state 1}\end{array}\,\middle|\,\begin{array}{c}\text{the site has state 0,}\\ \text{a neighbour has state 1}\end{array}\right)\\
&= P(\text{switch})\frac{z-1}{z}P\left(\begin{array}{c}\text{the picked,unknown}\\ \text{neighbour has state 1}\end{array}\,\middle|\,\text{the site has state 0}\right)+P(\text{switch})\frac{1}{z}\cdot 1
\end{aligned}$$

can be calculated which leads to

$$P((0,1)\to(1,1)) = \lambda\cdot dt\frac{z-1}{z}\frac{X_{01}}{X_{00}+X_{01}}+\lambda\cdot dt\frac{1}{z}. \tag{5.62}$$

Hereby the chance to pick the known neighbour with state 1 has a positive influence. As $X_{10}=X_{01}$ also

$$P((1,0)\to(1,1)) = P((0,1)\to(1,1)),\quad P((0,0)\to(0,1)) = P((0,0)\to(1,0)), \tag{5.63}$$

which concludes the calculation of the transition probabilities. We finally apply the mean-field theorem Corollary 5.1 and get

$$\begin{pmatrix}\varphi_{00}\\ \varphi_{01}\\ \varphi_{10}\\ \varphi_{11}\end{pmatrix}' = \begin{pmatrix}-\varphi_{00}\left(P((0,0)\to(0,1))+P((0,0)\to(1,0))\right)\\ \varphi_{00}P((0,0)\to(1,0))-\varphi_{10}P((1,0)\to(1,1))\\ \varphi_{00}P((0,0)\to(0,1))-\varphi_{01}P((0,1)\to(1,1))\\ \varphi_{01}P((0,1)\to(1,1))+\varphi_{10}P((1,0)\to(1,1))\end{pmatrix}.$$

Clearly, the third differential equation is redundant as $\varphi_{10} = \varphi_{01}$. We will only use φ_{10} henceforth. Furthermore, inserting (5.61) and (5.62), formula manipulations lead to the closed ODE system

$$\begin{pmatrix}\varphi_{00}\\ \varphi_{10}\\ \varphi_{11}\end{pmatrix}' = \begin{pmatrix}-2\lambda\cdot dt\varphi_{00}\frac{z-1}{z}\frac{\varphi_{10}}{\varphi_{00}+\varphi_{10}}\\ \lambda\cdot dt\varphi_{10}\frac{z\varphi_{00}-z\varphi_{10}-2\varphi_{00}}{z(\varphi_{00}+\varphi_{10})}\\ 2\lambda\cdot dt\varphi_{10}\frac{z\varphi_{10}+\varphi_{00}}{z(\varphi_{10}+\varphi_{00})}\end{pmatrix}.$$

Also the third equation for φ_{00} is redundant, as $\varphi_{11}+2\varphi_{10}+\varphi_{00} = zN$. It will not be considered henceforth. As $\varphi_1 = \frac{1}{z}(\varphi_{11}+\varphi_{10}), \varphi_1' = \frac{1}{z}(\varphi_{11}'+\varphi_{10}')$ we may add the first two differential equations to gain a differential equation for φ_1. This is used to replace the differential equation for φ_{11}. A lot of terms vanish after formula manipulation. We gain

$$\begin{pmatrix}\varphi_1\\ \varphi_{10}\end{pmatrix}' = \begin{pmatrix}\lambda\cdot dt\frac{\varphi_{10}}{z}\\ \lambda\cdot dt\varphi_{10}\left(\frac{z-1}{z}\left(1-2\frac{\varphi_{10}}{z(N-\varphi_1)}\right)-\frac{1}{z}\right)\end{pmatrix}. \tag{5.64}$$

Hereby the initial conditions

$$\varphi_1(0) = X_0, \quad \varphi_{10}(0) = zX_0\frac{N-X_0}{N} \tag{5.65}$$

hold due to uniformity of the initial distribution.

Unfortunately the gained differential equation system is numerically awfully stated. As X_1 approaches N when the population grows, the divisor in the second equation approaches 0 with similar speed causing heavy problems for any numerical ODE solver. Hence, we substitute:

$$\varphi := \varphi_1, \quad \psi := \frac{X_{10}}{z(N-X_1)}$$

and receive the much more comfortable system

$$\begin{pmatrix}\varphi\\ \psi\end{pmatrix}' = \begin{pmatrix}\lambda\cdot dt\psi(N-\varphi)\\ \lambda\cdot dt\psi(1-\psi)\frac{z-2}{z}\end{pmatrix}, \quad \varphi(0) = \psi(0) = X_0. \tag{5.66}$$

Also the choice of ψ can be interpreted. Its behaviour depicts the temporal development of the conditional probability

$$P(\text{pick a neighbour with state 1}|\text{the regarded site is state 0}).$$

We analyse (5.66). Our first observation shows, that the approximation works perfectly for $z = 2$ – i.e. the symmetric, two sided 1-d neighbourhood. Here $\psi = \psi(0) = X_0$ leads to a perfect match with the linear ODE (5.50) derived for the two-sided 1-d neighbourhood. For $z = 1$ the equation cannot be applied as the transformations from (5.64) to (5.66) required the division by $z-1$. Nevertheless, the prior stated ODE system (5.64) for $z = 1$ precisely leads the stated linear growth model for the one-sided neighbourhood.

The second observation deals with the limit case $z \to \infty$. Here, the formula leads to the standard Mean-field approximation of the model, the logistic growth as $\frac{z-2}{z} \underset{z\to\infty}{\longrightarrow} 1$.

Due to its simple shape, we attempt to simplify equation (5.66) even further. Dividing the two ODEs with ψ interpreted as a function of φ we gain

$$\frac{d\psi(\varphi)}{d\varphi} = \frac{1-\psi(\varphi)}{1-\varphi}\left(\frac{z-2}{z}\right)$$

which can be solved by separation to

$$\left(\frac{1-\psi(\varphi)}{1-X_0}\right)^z = \left(\frac{1-\varphi}{1-X_0}\right)^{z-2} \Rightarrow (1-\psi(\varphi)) = (1-\varphi)^{\frac{z-2}{z}}(1-X_0)^{\frac{2}{z}}.$$

Thus, we can write

$$\varphi' = \lambda \cdot dt\psi(1-\varphi) = \lambda \cdot dt\left(1-(1-\varphi)^{\frac{z-2}{z}}(1-X_0)^{\frac{2}{z}}\right)(1-\varphi), \quad \varphi(0) = X_0 \tag{5.67}$$

which finally leads to one closed first order differential equation comparable with the approximations (5.50) and (5.40). Its solution is expected to be a mean-field model for any **symmetric** neighbourhood with z elements.

The crucial second factor of the equation is clearly

$$\overline{\varphi} := 1-(1-\varphi)^{\frac{z-2}{z}}(1-X_0)^{\frac{2}{z}}$$

which is close to φ for large values of z and is close to X_0 for small values of z. In some ordering sense (dependency of φ) we get

$$X_0 \preceq \overline{\varphi}(t) \preceq \varphi(t).$$

Unfortunately we cannot expect that the number of neighboured sites z is sufficient information to accurately predict the aggregated numbers of Scenarios a, b and c. Figure 5.17 shows results of the pair approximation compared with three different variants of Scenario c, the Von Neumann neighbourhood. In each of the used neighbourhood scenarios, a site has $z = 4$ neighbours and the neighbourhood is symmetric. We investigate

- **Scenario c.1** a standard Von Neumann neighbourhood,

$$S(c_{i,j}) := \{c_{k,l} : d_1(c_{i,j}, c_{k,l}) \leq 1\} = \{c_{i-1,j}, c_{i,j-1}, c_{i+1,j}, c_{i,j+1}\},$$

- **Scenario c.2** a two sided 1-d neighbourhood, wherein each site's neighbours are the two sites to the left and to the right,

$$S(c_{i,j}) := \{c_{i-2,j}, c_{i-1,j}, c_{i+1,j}, c_{i+2,j}\},$$

  and finally

- **Scenario c.3** a two sided 1-d neighbourhood, wherein each site's neighbours are the 13th and 37th (some randomly chosen prime numbers) site to the left and right,

$$S(c_{i,j}) := \{c_{i-37,j}, c_{i-13,j}, c_{i+13,j}, c_{i+37,j}\}.$$

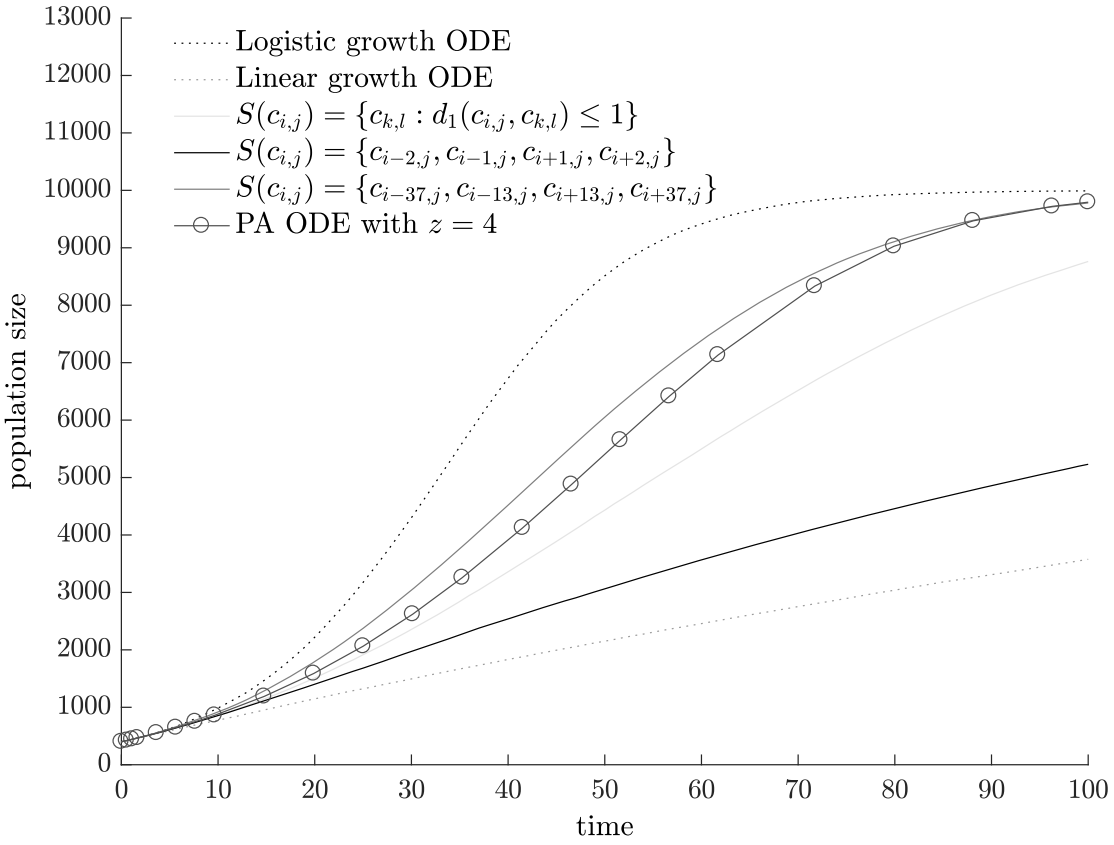


Figure 5.17: Solid lines show model scenarios with different neighbourhoods. Although all of them use a neighbourhood with four sites, the results differ a lot. The pair approximation is sketched with circles while the dotted lines show the limit cases for the logistic growth and the linear growth ODE models.

The three sub-scenarios lead to completely different results for the aggregate number of the microscopic model, though the PA ODE (5.67) for $z = 4$ states only one solution curve. The mean-field model somehow suits in the sense that the solution curves lie somewhere in the middle between the logistic growth ODE and the linear growth ODE. However we simply cannot expect good approximation results if geometry of the neighbourhood is ignored.

This problem is well known and several attempts to find specific constants to achieve convergence were made to solve it. In literature one finds concepts like the correlated pair approximation (CPA, see [Benoit et al., 2006, Berg et al., 1998]) or the improved pair approximation (IPA, see [Satô et al., 1994]).

It is clear that artificially increasing the state-space of the microscopic sub-models even more, e.g. by defining not only pairs but triplets or even the total neighbourhood as individuals of the model, also the geometry of the neighbourhood can be included to the mean-field

approximation and the resulting numeric solutions would suit better. We attempted a strategy with $|\Gamma'| = 10$ different states (compared to $|\Gamma'| = 4$ for the pair-approximation). The numeric solutions of the resulting system of 10 coupled non-linear differential equations are quite a good fit to the microscopic model results, but as

- the resulting ODE systems are huge and confusing,
- their derivation is, equally, long, confusing and requires tricky probabilistic calculations, and
- the quality of the fit is not good enough to justify the exhausting derivation process

we decided not to include the results in this work.

Summarising, local interactions of sub-models as presented on the example of different neighbourhood concepts on a lattice are still a huge challenge for mean-field approximation. Hereby a trade-off between reasonable complexity of the mean-field model especially regarding size of the state-space and validity of the mean-field model can be observed. In all our attempts to find a useful mean-field model to a microscopic simulation model with real local interactions, the additional effort of deriving a pair- or even higher order- approximation hardly paid off in comparison to the efforts of the derivation process. More reasonable results could be achieved by artificially adjust classic mean-field approximations like the logistic growth ODE by modifying parameters or by adding damping terms, respectively, to depict the slower population growth rate for smaller z with $2 < z < N$.

5.4 Mean-Field Analysis for Microscopic Models with High Sensitivity

In the language of modelling and simulation the broadly known term **sensitivity** denotes that a given system (real or modelled) reacts massively to small disturbances. This term has usually a negative connotation as applying a *sensitive model* for actual quantitative decision-support requires, that the model parameters need to be measured/calculated very carefully to receive valid model outcome. Mathematically spoken, a model is called sensitive if two slightly different (input-) parameters i_1, i_2 lead to heavily differing outcomes o_1, o_2:

$$\|i_1 - i_2\| \ll \|o_1 - o_2\| . \tag{5.68}$$

One of the most famous examples for a sensitive system is the weather. The fact that valid forecasting of weather via models, especially for medium- and long- range time periods requires extensive and precise data collection, has already been known in the early 1980s. Weather and therefore also the used weather-forecasting models react very sensitively on missing or slightly flawed input data [Bengtsson et al., 1984]. Especially the required amount of collected data for parametrisation of the forecasting models exponentially increases with the desired range of forecast.

Anyway, the level of sensitivity of any model needs to be investigated before application to avoid misinterpretation of the results. For differential equation models there are analytical methods to precisely detect and quantify sensitivity in model parameters. For most other modelling approaches, especially microscopic models, these methods are missing. In most cases the only option to detect sensitive parameters is the trial-and-error approach: *parameter sweep*. In fact for stochastic microscopic models even this method is not really sufficient.

As mentioned in Section 2.3.1 one of the key features of microscopic models is the ability to depict behaviour which is at least partly unpredictable, i.e. emergent. Some of the best models to depict this kind of behaviour are found investigating cellular automata which, though defined via simple rules depict very complex behaviour. We will use a case study to show two possible ways, how sensitivity might occur in microscopic models and, at least for the second case, how it might be detectable even **without executing the model**.

5.4.1 Sensitivity Case Study: Game of Life

As mentioned in Section 2.1.2, **Game of Life (GoL)** by John Horton Conway was probably the first cellular automaton (CA) which has become prominent to the broad public as it was introduced as a *mathematical game* and not as a numb piece of theoretical research.

Model 4.1: Game of Life (original by J.H. Conway)

With regards to the given definition of a cellular automaton (see Definition 1.1) the classic rules of Conways version of GoL can be given as follows.
Initial Setup:

- We investigate a rectangular grid of $N_1 \times N_2 =: N$ sites (in the original version, an infinite grid was used as gedankenexperiment). We denote

$$C := \{c_{i,j},\ i \in \{0, \dots, N_1\},\ j \in \{0, \dots, N_1\}\}$$

  as the site-space (already combined with the correct two-dimensional index mapping) and denote $c_{i,j}(t)$ as the site's time-dependent state. Each site's state is, analogous to the growth model in the last section, either 0 (dead) or 1 (alive).

Dynamics:

- The Moore neighbourhood is used:

$$S(c_{i,j}) = \{c_{k,l} : d_\infty(c_{i,j}, c_{k,l}) \leq 1\}.$$

  Hereby neighbourhoods that overlap with the grid boundaries are cut.

- The cellular automaton is updated as follows:

$$c_{i,j}(t) = 0 \to c_{i,j}(t+1) = 1 \Leftrightarrow |\{c_{k,l} \in S(c_{i,j}) : c_{k,l}(t) = 1\}| = 3 \quad (5.69)$$
$$c_{i,j}(t) = 0 \to c_{i,j}(t+1) = 0 \Leftrightarrow |\{c_{k,l} \in S(c_{i,j}) : c_{k,l}(t) = 1\}| \neq 3 \quad (5.70)$$
$$c_{i,j}(t) = 1 \to c_{i,j}(t+1) = 1 \Leftrightarrow |\{c_{k,l} \in S(c_{i,j}) : c_{k,l}(t) = 1\}| \in \{2,3\} \quad (5.71)$$
$$c_{i,j}(t) = 1 \to c_{i,j}(t+1) = 0 \Leftrightarrow |\{c_{k,l} \in S(c_{i,j}) : c_{k,l}(t) = 1\}| \notin \{2,3\} \quad (5.72)$$

These rules are usually understood better, when presented in textual form: A living site stays alive or survives, respectively, the regarded time-step if it has two or three living neighbours – otherwise it dies. A dead site becomes alive if it has exactly three living neighbours – otherwise it remains dead. One can imagine these rules to emerge from a never-ending civilisation process: On the one hand, no one wants to live in an overcrowded area or in a lonely area. Hence, an individual leaves a certain area (site) if it gets overcrowded (i.e. > 3 neighbours) or if the neighbourhood dies out (i.e. < 2 neighbours). On the other hand, an individual only settles on a specific spot if and only if the conditions are perfect (i.e. there are exactly 3 living neighbours).

The presented definition is often used as a paragon for a classic cellular automaton as it is

- easy to understand,
- easy to describe, but not trivial,
- cleanly deterministic,
- easily visualised,
- is known to have a nearly unpredictable temporal behaviour.

Especially for its last feature Conway's Game of Life has become famous. A lot of interesting patterns, i.e. characteristic formations of living sites, have been found via lots of (basically) heuristic research during the last decades. Some of them are known to remain static, others are known to have an oscillating behaviour. So-called gliders are known to "fly" across the site-space and would carry on flying forever if the site-space would not be finite. A snapshot showing some classic patterns is found in Figure 5.18. Since the development of the model in 1970 discovery of new interesting patterns has become a sportive activity for thousands of scientists.

Regarding specific patterns the GoL turns out to be massively sensitive on changes in the initial condition. To see this we randomly created a reference initial condition: 20 percent of a $50 \times 50 = 2500$ site grid, i.e. 500 sites were randomly (uniformly) chosen to be alive at the beginning of the simulation while the remaining sites are defined as inactive. To analyse sensitivity,

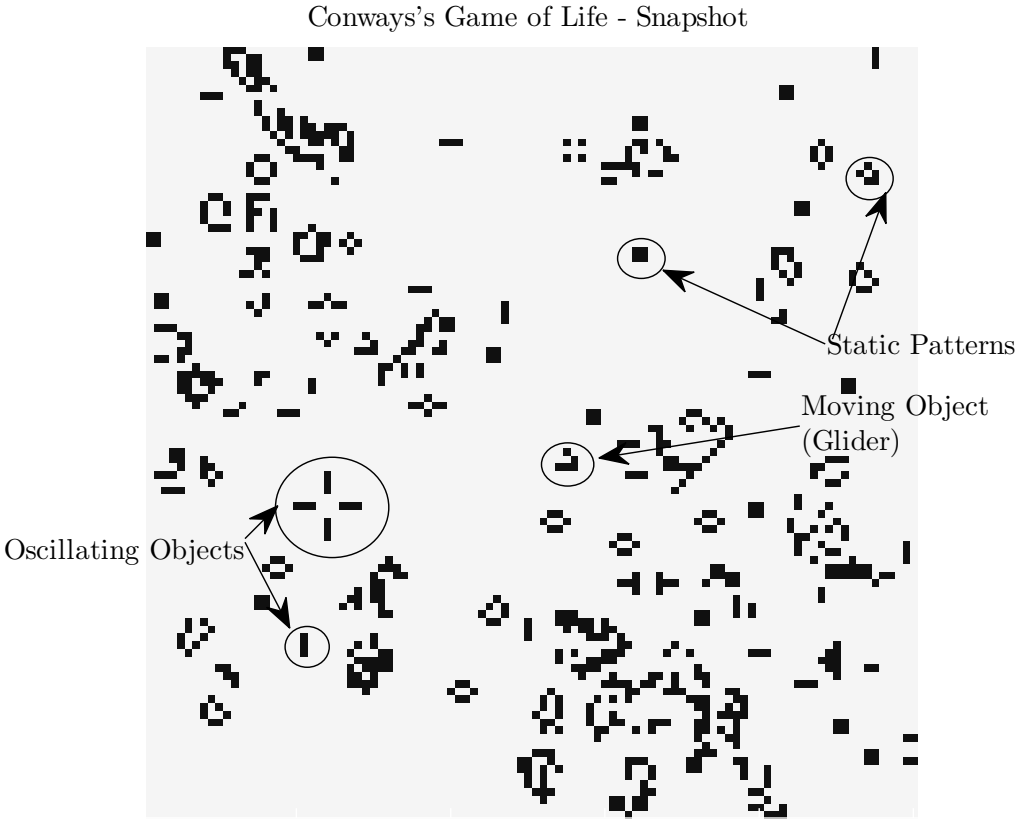


Figure 5.18: Some classic Conway's Game of Life patterns

one additional living site was systematically added to the reference initial condition shown in the upper-middle of Figure 5.20. I.e. the initial condition was changed by $1 - \frac{501}{500} = 0.002 = 0.2$ percent. Afterwards, 100 steps of the GoL were executed with both initial conditions, the reference and the modified one. After 100 steps both grids of the GoL are compared again and their difference was evaluated as the total number of states that have a different value. I.e. given the reference initial condition C_0^{ref} and the modified initial condition C_0^{mod}, both interpreted as a $N_1 \times N_2$ matrix, their difference is calculated via

$$\left\| C_1^{\text{ref}}00 - C_1^{\text{mod}}00 \right\|_F ,$$

with the Frobenius norm F. Note, that always

$$\left\| C_0^{\text{ref}} - C_0^{\text{mod}} \right\|_F = 1$$

as only one additional living site is added. The results of this experiment are shown in Figure 5.19 plotted as a histogram. Of all 2000 inactive sites there were almost 500 cases in which the

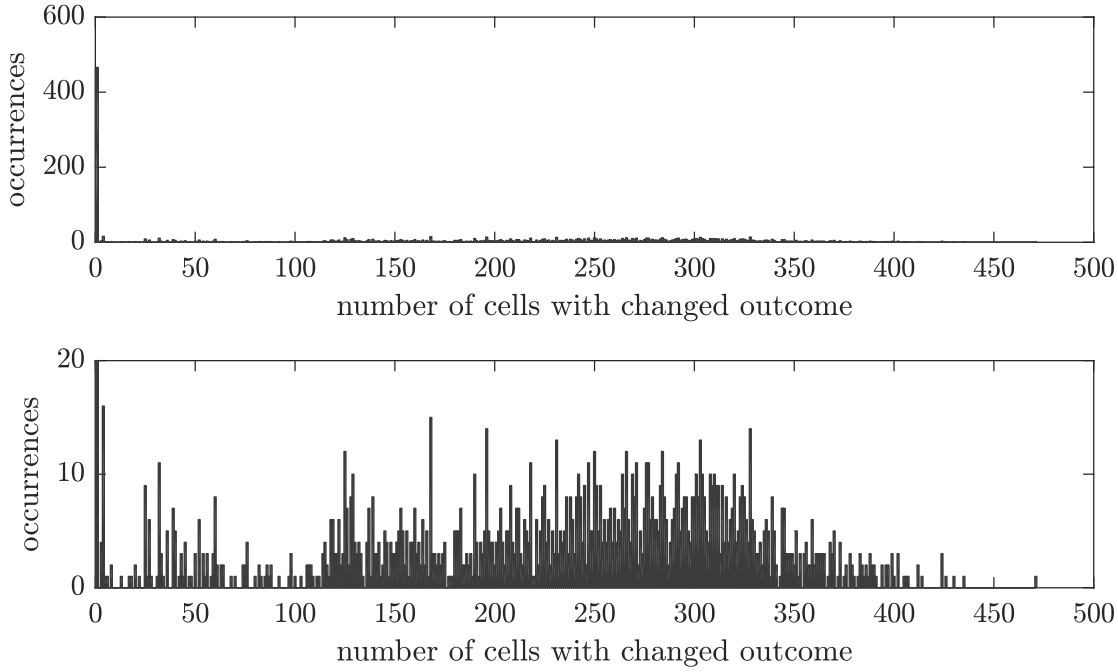


Figure 5.19: The two plots show the same histogram of the 2000 described experiments, wherein the lower plot is equivalent to the upper one with restricted y-axis. In each of those experiments one additional living site was added to a reference initial condition of the GoL. Moreover, the final states of the modified initial conditions after 100 GoL steps were compared to the final state of the reference initial condition. The outcome of each experiment is the number of sites that have a different state.

change of the site from dead/inactive to alive/active had **no** influence on the simulation result after 100 steps whatsoever. This is especially prominent in the upper plot of the figure. Yet, there are also specific spots on the grid which have a huge influence on the simulation outcome. In most cases they lead to a change of 200 to 350 sites compared to the reference result C^{ref}_{100}, but also higher values are possible. Changes of $1-100$ sites are comparably rare. This is best seen in the lower part of the plot which is, essentially, a zoom of the upper plot. A representative picture is given in Figure 5.20. Hereby the reference initial condition is shown in the upper-middle plot while the corresponding model outcome after 100 GoL steps is shown below. The left part of the figure shows a modification of a site which had no influence on the simulation result, the right part of the image shows a modification of the initial condition that changed the state of 414 sites after 100 steps.

Summarising, the **pattern structure of the GoL** can be seen to be **very sensitive** with respect to its **initial condition**. Finding an explanation for this is very complex and can hardly be generalised. Yet, compared with similar models, the feature of a solely deterministic update function seems to be responsible for high sensitivity while stochastic update rules often lead to

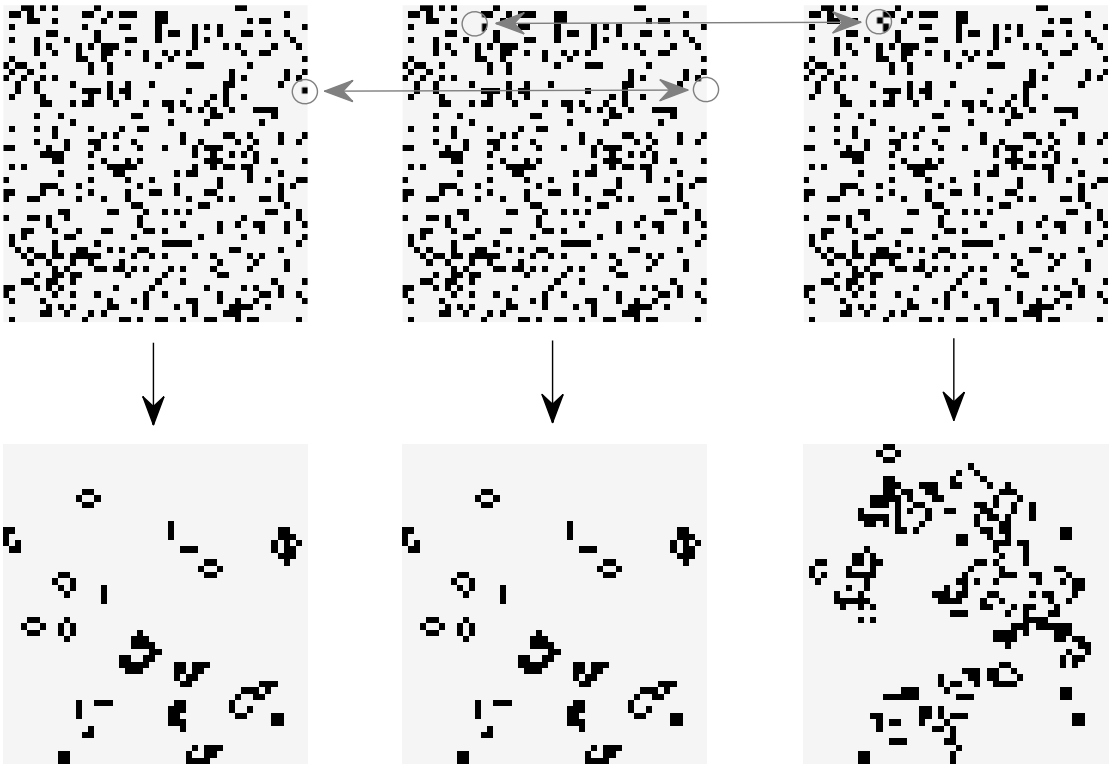


Figure 5.20: Based on a reference initial conditions seen in the upper-middle, two additional living sites were added leading to the initial conditions to the upper-right and the upper-left. 100 update-steps of the GoL CA were simulated and the results are plotted directly below the corresponding initial condition. Only the living site added in the right initial condition had an impact on the outcome.

more stable simulations. This might sound unintuitive at first, but can be verified by experiments – the shuffled GoL model introduced in Section 5.4.3 is a perfect example how the introduction of stochastic elements reduces the sensitivity.

5.4.2 Failed Mean-Field Analysis Case Study: Game of Life

Based on the observed sensitivity problem it is quite clear that the pattern resulting from the Game of Life (GoL) cannot be obtained with any macroscopic model. There is simply no macroscopic approach that reacts that swiftly on minimal changes in the initial condition, especially not a differential equation. As the total number of changed sites also has an influence on the aggregated numbers, i.e. the total number of living and dead sites, it may also be doubted that the mean-field can be modelled validly with a mean-field model.

Yet, we attempt to perform a mean-field analysis like in the previous sections.

At this point, it is important to mention that it is not usual to analyse the Game of Life (GoL) on the aggregate level as it is not intended to be. Yet, we want to gain insights into the behaviour of the system and, in general, into the effects of deterministic local interactions.

We follow the step-by-step process defined for mean-field analysis in Section 4.6.2:

- First, we find, that there are $N_1 \cdot N_2 =: N$ individual sub-models in the cellular automaton, namely each site in the site-space.
- The common state-space of the sites is found by $\Gamma := \{0, 1\}$, i.e. inactive/dead and active/alive. Let moreover

$$X_1(t) := \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbb{1}_1(c_{i,j}(t)), \quad X_0(t) := N - X_1(t).$$

- Sites may switch from 0 to 1 as well as from 1 to 0.
- The corresponding transition-probabilities could be calculated as follows: We assume that living and dead sites are uniformly spread among the grid (which is, of course, only an approximation). A dead site becomes alive if precisely 3 of its 8 neighbours (Moore neighbourhood) is alive. According to combinatorics, there are $\binom{8}{3}$ possible alignments of 3 among 8 sites. The chance for each of these alignments can be approximated using the total fraction

$$x := \frac{X_1(t)}{N}$$

of alive sites via $x^3(1-x)^5$. Hence,

$$P(0 \to 1) \approx \binom{8}{3} x^3(1-x)^5 = \binom{8}{3} \frac{X_1(t)^3(N - X_1(t))^5}{N^8}. \tag{5.73}$$

A site dies if its neighbour-count is smaller than 2 or larger than 3. We get a valid estimation for the corresponding probability when approximating the probability for the inverse event and subtracting this number from one

$$P(1 \to 0) \approx 1 - \left(\binom{8}{2} \frac{X_1(t)^2(N - X_1(t))^6}{N^8} + \binom{8}{3} \frac{X_1(t)^3(N - X_1(t))^5}{N^8} \right). \tag{5.74}$$

Equipped with the correct transition probabilities we can set up the mean-field differential-equation model and get (after formula manipulations)

$$\frac{d}{dt}\varphi_1 = \binom{8}{3} \frac{\varphi_1(t)^3(N - \varphi_1(t))^5}{N^7} + \binom{8}{2} \frac{\varphi_1(t)^3(N - \varphi_1(t))^6}{N^8} - \varphi_1 \tag{5.75}$$

$$\varphi_1(0) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \mathbb{1}_1(c_{i,j}(0)) \tag{5.76}$$

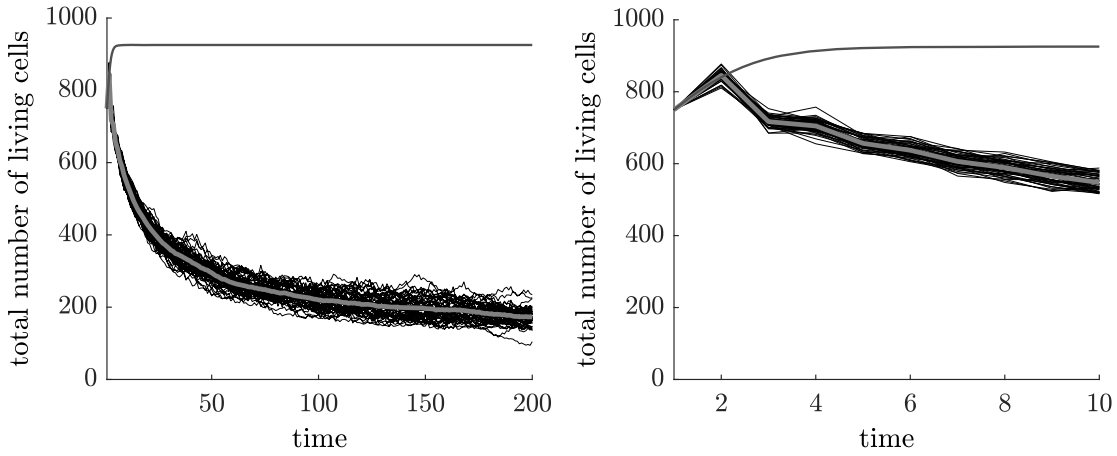


Figure 5.21: Fifty GoL results with a 2500 sites automaton and randomly chosen 1000 initially alive sites (black), compared with their mean-value (red) and the numeric solution of the Mean-Field ODE (blue). The right plots show a zoom to the first 10 time-steps.

Note, that we used $\varphi_0 + \varphi_1 = N$ to simplify the equations.

As expected the established mean-field analysis is a great failure as the resulting ODE system cannot depict the temporal behaviour of the GoL, not even when we randomly permute the initial condition matrix and take the mean value. Figures 5.21 and 5.22 show how the derived mean-value solution curve of (5.75) drifts off the microscopic GoL results – they do not even match qualitatively. The great fluctuations seen in Figure 5.22 for the GoL results additionally underline that a correct macroscopic description of the dynamics of the aggregated numbers of this cellular automaton is very likely not possible at all.

More about this can be found in [Bicher and Popper, 2015]. Also Figure 5.21 was already published there.

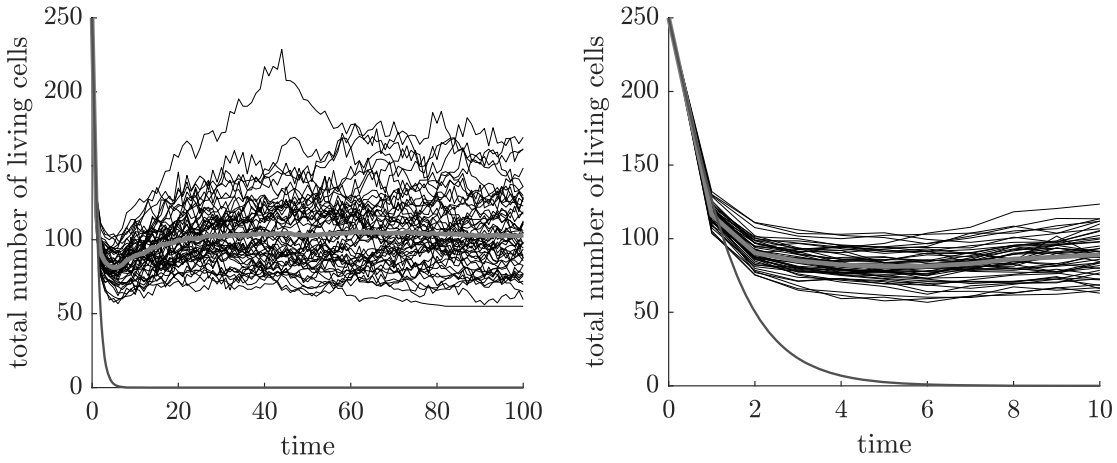


Figure 5.22: Fifty GoL results with a 2500 sites automaton and randomly chosen 250 initially alive sites (black), compared with their mean-value (red) and the numeric solution of the Mean-Field ODE (blue). The left plots show a zoom to the first 10 time-steps.

5.4.3 Sensitivity Case Study: Shuffled Game of Life

The results of the last section showed that the Game of Life (GoL) is a nice example for a microscopic model that cannot be analysed with mean-field methods, at least not with reasonable efforts. While the question, how to find a reasonable mean-field model for the GoL remains unanswered, it is possible to investigate the problem from a different angle:

> *Which parts of the microscopic model, i.e. the cellular automaton (CA), do we have to change or add, so that the mean-field model (5.75) suits?*

As mentioned in the last section the mean-field model is based on the assumption that living and dead (inactive) sites are uniformly spread among the grid, which is reflected in the local neighbourhood of any site. As seen in the results this assumption is clearly wrong as the Game of Life is very sensitive with respect to specific pattern formations. Hence, to establish a model that suits the mean-field model, it is necessary to bring some additional mixing into the model. We do that following [Bicher and Popper, 2015] by adding an additional "shuffling" phase in addition to the update phase.

Model 4.2: Shuffled Game of Life

With respect to the given definition of a cellular automaton (see Definition 1.1) the rules of the shuffled version of the GoL can be stated as follows.

Initial Setup:

- We investigate a rectangular grid of $N_1 \times N_2 =: N$ sites. We denote

$$C = \{c_{i,j},\ i \in \{0, \dots, N_1\},\ j \in \{0, \dots, N_2\}\}$$

  as the site-space (already combined with the correct two-dimensional index mapping) and refer to $c_{i,j}(t)$ as the site's time-dependent state.

- Each site's state is either 0 (dead) or 1 (alive).

Dynamics:

- The Moore neighbourhood is used:

$$S(c_{i,j}) = \{c_{k,l} : d_\infty(c_{i,j}, c_{k,l}) \leq 1\}.$$

  Hereby neighbourhoods are toroidally extended beyond the CA border.

- The cellular automaton is updated as follows:

$$c_{i,j}(t) = 0 \to c_{i,j}(t+1) = 1 \Leftrightarrow |S(c_{i,j}) = 1| = 3 \tag{5.77}$$
$$c_{i,j}(t) = 0 \to c_{i,j}(t+1) = 0 \Leftrightarrow |S(c_{i,j}) = 1| \neq 3 \tag{5.78}$$
$$c_{i,j}(t) = 1 \to c_{i,j}(t+1) = 1 \Leftrightarrow |S(c_{i,j}) = 1| \in \{2, 3\} \tag{5.79}$$
$$c_{i,j}(t) = 1 \to c_{i,j}(t+1) = 0 \Leftrightarrow |S(c_{i,j}) = 1| \notin \{2, 3\} \tag{5.80}$$

  Afterwards, every site **switches state with a randomly chosen second site**.

Remark 4.1:

To be precise, by addition of this mixing phase the microscopic model defined above is not a CA anymore, but can be interpreted as an agent-based model on a lattice. The influence range of a site exceeds, by far, the defined Moore neighbourhood.

Nevertheless, the case study in this section also works nicely if, instead of the stochastic shuffling process, deterministic movement rules similar to but slightly more rigorous than in the SIR lattice model 2.1 are defined. Hereby we would regain the interpretation of the model as a CA, as the site-states are deterministically shifted according to their index on the grid or, to be precise, the remainder of their x and y position on the grid by four.

As they are more difficult to define and the validity of the mean-field approximation is not that obvious we did not include them in above model and stick with the stochastic mixing.

It is clear that the shuffling rule leads to a uniform spread of alive and dead sites. Hence, also the mean-field approximation holds, which can be seen in the scenario for $X_0 = 750$ in Figure 5.23. Although this toy-model is based on the same rules as the classic Game of Life formulation, completely different results and asymptotic behaviour are achieved as obviously the development of patterns and clusters is the most crucial part of Conway's model. Nevertheless, the newly defined model still reveals some interesting features regarding a different kind of stability of stochastic microscopic simulation.

We choose an initial value of $X_0 = 500$ sites, which is 20 percent of total 2500 sites, and calculated the sample mean and variance of 50 simulation-runs. Hereby a very high variance (standard deviation) raises interests. Closer analysis of this result shows that about two third of the simulation-runs converge towards about 920 sites, as they did in the first scenario with $X_0 = 750$ (in Figure 5.23), but in about one third of the simulation-runs the population dies out. Consequently, the **sample-mean**, which converges to about 600 sites, **is not a representative** output measure for the behaviour of the simulation results! This is illustrated in Figure 5.24.

This on the first view unpredictable behaviour can be explained analysing the mean-field ODE (5.75) or, to be precise, the steady-states of it. It turns out, that the ODE has exactly three positive steady states between 0 and N of which two (the lowest, of course $X(t) = 0$, and the highest, about $X(t) \approx 0.37N$) are strongly attracting and the one in the middle ($X(0) \approx 0.192N$) is strongly repelling (Figure 5.25). Hence, the initial value $X_0 = 750 = 0.2N$ is extremely sensitive regarding fluctuations. Thus, experimental results need to be interpreted differently.

This case study, once again, reveals the necessity of a mathematical base for stochastic microscopic models. Without assistance of the ODE formulation and the related steady state analysis of the differential equation (in combination with Ljapunov stability theory) the strange splitting of the resulting curves would have remained unexplained. Using a sample mean value(red curve in Figure 5.24) as base for any kind decision-support can lead to horrible misplannings.

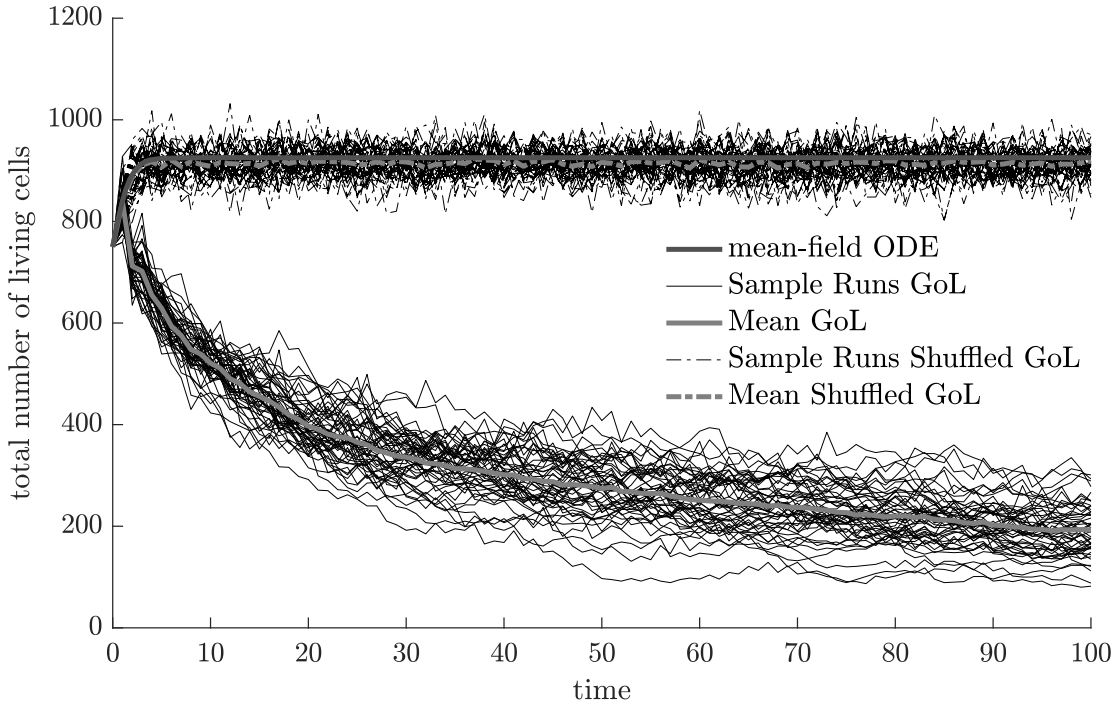


Figure 5.23: Comparison between the results of the classic GoL, the shuffled GoL and the calculated ODE mean-field model for 30 percent initially living sites, i.e. $X_0 = 750 = 0.3 \cdot 2500$.

Summarising, two different types of sensitivity have been investigated in the course of the Game of Life case study:

- local sensitivity and
- mean-field sensitivity.

The second source can be predicted and analysed by investigating the stability of the mean-field model. The first one is usually unpredictable and very difficult to deal with. We do not claim that these two are the only sources for sensitivity in microscopic models, yet we could, so far, attribute any sensitivity issue in investigated microscopic models to one (or both) of these two. Of course, it is in the first place impossible to decide, which of the two types occurred when the microscopic model results seem unusual, nevertheless, mean-field analysis can be used to either exclude or to analyse the second one.

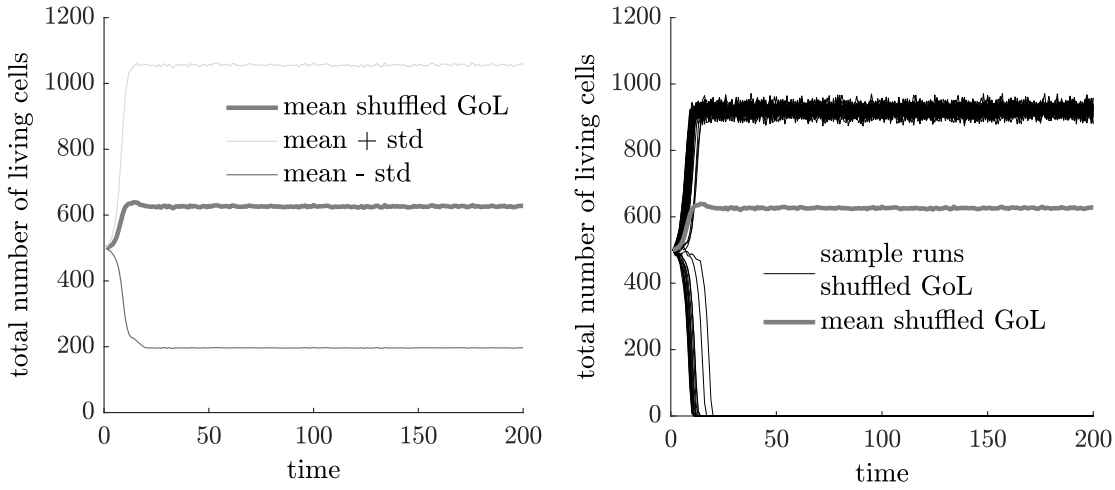


Figure 5.24: Interestingly high standard deviation for initial condition $X_0 = 500$ sites as seen in the left plot results from strange instability of the simulation results as seen in the right plot. About two third of the simulation results lead to an approximately stable population while it quickly dies out in all other runs.

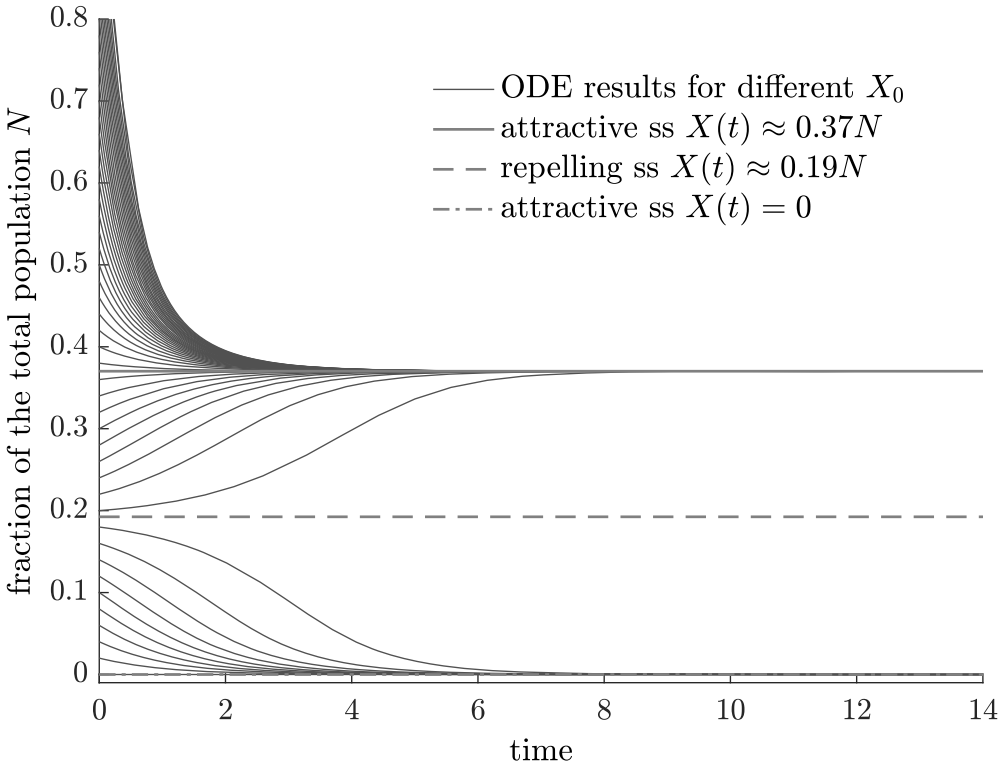


Figure 5.25: Steady state (ss) analysis of the mean-field ODE (5.75).

5.5 Mean-Field Analysis for State-Continuous Models

As they are considerably easier to develop, most microscopic models are state-discrete, i.e. the states each individual sub-model can be found to be member of a discrete and usually finite set. Yet, in some cases the inclusion of state continuous variables is unavoidable, e.g. because a discretisation process would involve too many errors or is unnecessarily complicated.

A very convincing example for this problem can be seen in [Glock et al., 2012] wherein (primarily) the discretisation of the basically continuous variable *age* in a system dynamics model for obesity required the use of hundreds of system dynamics stocks. Note, that they all had to be drawn single-handedly on the AnyLogic graphical user interface to finally execute the model.

On the contrary, to many other modelling approaches continuous variables in microscopic approaches **must not** necessarily **be discretised** to execute the model. An example for this is given in the following section. We perform mean-field analysis of a state-continuous random-walk model.

5.5.1 Case Study: Random Walk Model for Groundwater Pollution

The case study in this section, a microscopic random walk model, was motivated from [Winkler et al., 2015] and [Winkler, 2014]. Herein random walk is used to simulate the spread of pollution in flowing groundwater. As this particle-based model was developed to pose for an approximation of the convection-diffusion equation, a candidate for a correct mean-field model is already known. In this section we want to apply the continuous-space mean-field theorem Corollary 5.3 to prove that this assumption is valid.

Model 5.1: Random Walk Model for Groundwater Pollution

We define the microscopic particle model via its initial setup and its dynamics.

Initial Setup:

- Let $N \in \mathbb{N}$ be a sufficiently large number of particles and define their time-dependent position by $I_i(t), i \in \{1, \dots, N\}$. It is found as a point on the two dimensional plane

$$I_i(t) \in \mathbb{R}^2.$$

- Initially, place all particles in the origin. I.e.

$$I_i(0) = (0,0)^T, \quad i \in \{1, \dots, N\}.$$

The model is updated in discrete time-steps of length dt.

Dynamics:

- Every time-step each particle is addressed once to change its position according to the following law.
- Define

$$v_i^1 := dt\vec{\alpha} := dt \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}, \quad \alpha_1, \alpha_2 \in \mathbb{R}, \tag{5.81}$$

and

$$v_i^2 := \sqrt{dt}\beta \begin{pmatrix} X \\ Y \end{pmatrix}, \quad \beta \in \mathbb{R}^+, \tag{5.82}$$

wherein X and Y are two standard normally distributed random variables. Moreover

$$I_i(t + dt) = I_i(t) + v_i^1 + v_i^2. \tag{5.83}$$

For the sake of giving a valid picture of the model, we denoted the approach as a *particle model* which is not stated as a separate modelling approach in Section 2.1. The reader might also interpret it as an agent-based approach without interaction, a microsimulation model or a dynamic Monte Carlo simulation.

Clearly, the movement of particles is split into two parts: a deterministic one, which corresponds to a shift into direction (α_1, α_2), and a stochastic one, which can be seen to spread the particles. For result plots the reader is referred to [Winkler et al., 2015] Section 5.

We will apply the Continuous Space MFT for Models with Static Population 5.3 directly:

1. Clearly, $\Gamma = \mathbb{R}^2$ and $d = 2$ as we investigate movement on a two-dimensional space. Hence, we need to find the transition kernel

$$\omega\left(t, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}\right) =: \omega\left(t, \vec{x}, \vec{y}\right) \tag{5.84}$$

that states, how likely a particle in state $(x_1, x_2)^T$ moves to $(y_1, y_2)^T$. According to the definition of the transition kernel Definition 4.10 we may use the finite difference approximation

$$\omega(t, \vec{x}, \vec{y}) \approx \frac{\tilde{P}(I_i(t + dt) = \vec{y} | I_i(t) = \vec{x})}{dt}$$

with the transition-density $\tilde{P}$. Thus, we need to calculate the density for the spatial movement of a model particle.

2. As a result of basic probability calculus a linear affine transformation of a standard Gaussian random number X leads to a Gaussian random number with shifted mean and scaled standard deviation via

$$X \sim \mathcal{N}_{0,1} \Rightarrow a + bX \sim \mathcal{N}_{a,b}.$$

This can be applied to both rows of the defined time evolution of each particle

$$I_i(t+dt) = I_i(t) + v_i^1 + v_i^2 = \left(I_i(t) + dt \begin{pmatrix}\alpha_1\\ \alpha_2\end{pmatrix}\right) + \sqrt{dt}\beta \begin{pmatrix}X\\ Y\end{pmatrix} \sim \begin{pmatrix}\mathcal{N}_{I_i(t)_1+dt\cdot\alpha_1,\sqrt{dt}\beta}\\ \mathcal{N}_{I_i(t)_2+dt\cdot\alpha_2,\sqrt{dt}\beta}\end{pmatrix}. \quad (5.85)$$

Hence, the transition density of the two-dimensional position of a particle is a product density of two independent Gaussian distributed random numbers with different mean but equivalent standard deviation. It follows that

$$\begin{aligned}\tilde{P}\left(I_i(t+dt) = \begin{pmatrix}y_1\\ y_2\end{pmatrix} \middle| I_i(t) = \begin{pmatrix}x_1\\ x_2\end{pmatrix}\right) &= \frac{1}{\sqrt{2\pi}\sqrt{dt}\beta}\exp\left(-\frac{(y_1-(x_1+dt\cdot\alpha_1))^2}{2dt\cdot\beta^2}\right)\\ &\cdot\frac{1}{\sqrt{2\pi}\sqrt{dt}\beta}\exp\left(-\frac{(y_2-(x_2+dt\cdot\alpha_2))^2}{2dt\cdot\beta^2}\right)\\ &= \frac{1}{2\pi dt\beta^2}\exp\left(-\frac{(y_1-x_1-dt\cdot\alpha_1)^2+(y_2-x_2-dt\cdot\alpha_2)^2}{2dt\cdot\beta^2}\right).\end{aligned}$$

and, as $\omega \approx \tilde{P}/dt$,

$$\omega\left(t, \begin{pmatrix}x_1\\ x_2\end{pmatrix}, \begin{pmatrix}y_1\\ y_2\end{pmatrix}\right) \approx \frac{1}{2\pi dt^2\beta^2}\exp\left(-\frac{\|\vec{y}-\vec{x}-dt\vec{\alpha}\|_2^2}{2dt\cdot\beta^2}\right). \quad (5.86)$$

It is not surprising that the transition kernel does not depend on any aggregated numbers as the particles in the microscopic model do not interact.

3. After definition of the transition kernel, we need to calculate the coefficient functions $F1, F2$ and $F3$. We get

$$F1(t,\vec{x}) = \int_{\mathbb{R}^d}\omega(t,\vec{x},\vec{y}) - \omega(t,\vec{y},\vec{x})d\vec{y} = \frac{1}{2\pi dt^2\beta^2}\int_{\mathbb{R}^2} e^{-\frac{\|\vec{y}-\vec{x}-dt\vec{\alpha}\|_2^2}{2dt\cdot\beta^2}} - e^{-\frac{\|\vec{x}-\vec{y}-dt\vec{\alpha}\|_2^2}{2dt\cdot\beta^2}} = 0. \quad (5.87)$$

The integral vanishes, as with every $\vec{x}-\vec{y}$ also $\vec{y}-\vec{x}$ lies in $\mathbb{R}^2$. Moreover,

$$\begin{aligned} F2_1(t,\vec{x}) &= \int_{\mathbb{R}^d} \omega(t,\vec{y},\vec{x})(y_1-x_1)d\vec{y} \\ &= \frac{1}{2\pi dt^2\beta^2}\int_{\mathbb{R}^2} e^{-\frac{\|\vec{x}-\vec{y}-dt\vec{\alpha}\|_2^2}{2dt\cdot\beta^2}}(y_1-x_1)d\vec{y} \\ &= \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{dt}^3\beta}\int_{\mathbb{R}} e^{-\frac{(x_1-y_1-dt\alpha_1)^2}{2dt\cdot\beta^2}}(y_1-x_1)dy_1}_{=-\alpha_1}\underbrace{\frac{1}{\sqrt{2\pi}\sqrt{dt}\beta}\int_{\mathbb{R}} e^{-\frac{(x_2-y_2-dt\alpha_2)^2}{2dt\cdot\beta^2}}dy_2}_{=1} \end{aligned} \tag{5.88}$$

can be calculated. For symmetry reasons,

$$F2_2(t,\vec{x},\vec{y}) = -\alpha_2. \tag{5.89}$$

It remains to deal with $F3$. Analogously to $F2$ we calculate

$$\begin{aligned} F3_{2,1}(t,\vec{x}) = F3_{1,2}(t,\vec{x}) &= \frac{1}{2}\int_{\mathbb{R}^d}\omega(t,\vec{y},\vec{x})(y_1-x_1)(y_2-x_2)d\vec{y} \\ &= \frac{1}{2}dt^2\underbrace{\frac{1}{\sqrt{2\pi}\sqrt{dt}^3\beta}\int_{\mathbb{R}} e^{-\frac{(x_1-y_1-dt\alpha_1)^2}{2dt\cdot\beta^2}}(y_1-x_1)dy_1}_{=-\alpha_1} \\ &\quad\cdot\underbrace{\frac{1}{\sqrt{2\pi}\sqrt{dt}^3\beta}\int_{\mathbb{R}} e^{-\frac{(x_2-y_2-dt\alpha_2)^2}{2dt\cdot\beta^2}}(y_2-x_2)dy_2}_{=-\alpha_2} \\ &= \frac{dt^2\alpha_1\alpha_2}{2}. \end{aligned} \tag{5.90}$$

which is negligibly small, if dt is small. For the two remaining parameter functions we get

$$\begin{aligned} &F3_{1,1}(t,\vec{x}) \\ &= \frac{1}{2}\underbrace{\frac{1}{\sqrt{2\pi}\sqrt{dt}^3\beta}\int_{\mathbb{R}} e^{-\frac{(x_1-y_1-dt\alpha_1)^2}{2dt\cdot\beta^2}}(y_1-x_1)^2dy_1}_{=\alpha_1^2dt+\beta^2}\underbrace{\frac{1}{\sqrt{2\pi}\sqrt{dt}\beta}\int_{\mathbb{R}} e^{-\frac{(x_2-y_2-dt\alpha_2)^2}{2dt\cdot\beta^2}}dy_2}_{=1} \\ &= \frac{1}{2}(\alpha_1^2dt+\beta^2) \end{aligned} \tag{5.91}$$

of which the first part is, also, negligibly small, if dt is small. Hence,

$$F3_{2,2} = F3_{1,1} \approx \frac{1}{2}\beta^2. \tag{5.92}$$

According to the mean-field equation we finally obtain a linear partial differential equation (PDE) of second order.

Model 5.2: Convection-Diffusion Model

The model is given by

$$\frac{\partial}{\partial t}\varphi(\vec{x},t) = -\vec{\alpha}\nabla_x\varphi(\vec{x},t) + \frac{\beta^2}{2}\Delta\varphi(\vec{x},t), \tag{5.93}$$

wherein ∇ denotes the Nabla operator, and Δ stands for the Laplacian.

It poses for a valid mean-field model for the microscopic particle approach if dt is chosen sufficiently small.

As the random-walk model is a valid microscopic representation of the convection-diffusion equation it can be used as some type of a microscopic solver for the macroscopic PDE system. The mean-field theorem states the equivalence of the two systems only with respect to some errors depending on the step-width and the number of particles N. Note, that the transition kernel does not depend on the aggregated numbers and, hence, no additional error source depending on $\nabla\omega$ is present. The error with respect to N can be reduced easily by simply increasing the number of particles. To avoid the error with respect to the time-step width it would clearly be better to rephrase the whole model in a discrete event manner similar to the population model in Section 5.1. Unfortunately, this is not straight forward, as the individual particle processes do not only need to be time-continuous, but also space-continuous.

Remark 5.1:

Surprisingly, convection-diffusion processes are often calculated using a lattice model [Fibich and Gibori, 2010] which, in principle, leads to a discrete state-space again. Also completely deterministic microscopic approaches can be found [Dembele,].

Remark 5.2:

The results in this section are usually gained using methods from stochastic differential equations and Itô integration, which, in this case, is a lot quicker than doing the calculations on the density function-level (more about these concepts see e.g. [Gardiner, 2009]). The similarity with stochastic differential equations is seen particularly well in the diffusion parameter $\beta(X,Y)$ in (5.82), wherein correct time-scaling of the requires multiplication with $\sqrt{dt}$ and not with dt.

The case study in this section showed mean-field analysis of a model with solely continuous states and, especially, without interaction between the sub-models. Hence, we will present a second case study of a more advanced microscopic model that provides these features. Yet, we will not perform the mean-field analysis process in details as it would be too long and confusing (even) for this thesis.

5.5.2 Case Study: Age-Dependent SIR Model

In this section we want to give an example of a mean-field analysis of a model wherein the microscopic sub-models interact on a *hybrid state-space*. As mentioned, we will not discuss the corresponding mean-field analysis in detail but only state the main results. This case study will afterwards pose for the perfect base, to point out another very important problem within the analysis of microscopic models.

We refer to an agent-based SIR epidemics model very similar to the one discussed in Section 5.2.1. Yet, an additional individual feature influences the behaviour of the agents: their age.

First, the feature that a person's age has a crucial influence on the capabilities of dealing with a disease are long known and self-explanatory: The human body simply reacts more heavily on most types of diseases when the infected person is either too old or too young. As a result, the rate on which a person recovers from a disease (naturally or under medication) tends to be higher, if the infected person has a vital age.

Second, any kind of contact necessary for spreading an infectious disease to other individuals strongly depends on the age of the individual and its contact partners. For classic airborne diseases like measles or influenza we expect that the probability for infectious contacts between individuals which are about the same age is larger, than between individuals with a higher age-difference. If the age-difference is one generation – i.e. about 25 to 35 years – we expect that this probability has a second peak due to parent-child relationships, and finally a third one after about 50 to 70 years due to contacts with grandparents. Also, the total amount of different social contacts can be seen to depend on age.

Around 1927 epidemiologists Kermack and McKendrick published the first modelling concepts that include age-specific ideas to model the spread of an **Susceptible-Infectious-Recovered (SIR)** epidemic [McKendrick, 1925, Kermack and McKendrick, 1927]. Herein the two scientists were not only responsible for establishing the famous SIR ODE model already stated in equations (5.20)-(5.22), but had many ideas that exceed this basic concept. In [Hoppensteadt, 1974, Dietz and Schenzle, 1985] a very interesting model based on the ideas of, especially, McKendrick can be found.

Model 5.3: McKendrick Equation

Let $N \in \mathbb{N}$ be sufficiently large and let

$$S_0(a),\ I_0(a),\ R_0(a) \tag{5.94}$$

be positive functions on $[0, \infty)$ with

$$\int_0^\infty S_0(a) + I_0(a) + R_0(a)da = N. \tag{5.95}$$

These functions hereby depict population-densities for susceptible (S), infected (I) and recovered (R) individuals with respect to their age a (in years). Hence

$$\int_{c_1}^{c_2} S_0(a)da$$

describes the total number of susceptible individuals between c_1 and c_2. Moreover, the following system of partial differential equations describes the temporal development of these densities with time:

$$\frac{\partial S}{\partial t}(a,t) = -\frac{\partial S}{\partial a}(a,t) - (\lambda(a,t) + \mu(a))S(a,t), \tag{5.96}$$

$$\frac{\partial I}{\partial t}(a,t) = -\frac{\partial I}{\partial a}(a,t) + \lambda(a,t)S(a,t) - (\overline{\mu}(a) + \gamma(a))I(a,t), \tag{5.97}$$

$$\frac{\partial R}{\partial t}(a,t) = -\frac{\partial R}{\partial a}(a,t) + \gamma(a)I(a,t) - \mu(a)R(a,t). \tag{5.98}$$

Hereby

$$S(a,0) = S_0(a), R(a,0) = R_0(a), R(a,0) = R_0(a), \tag{5.99}$$

and

$$\nabla S(0,t) = \nabla I(0,t) = \nabla R(0,t) = 0. \tag{5.100}$$

Functions $\mu, \overline{\mu}$ and γ are positive, age-dependent death and recovery functions and

$$\lambda(a,t) := \int_0^\infty \kappa(a,b)\frac{I(b,t)}{N}db$$

poses for an infection-rate with contact kernel κ. Explanation and interpretation of the parameter-functions $\lambda, \gamma, \mu, \overline{\mu}$ and κ is given below.

As the state variable I appears in the definition of λ the partial differential equation (PDE) is non-linear. As also an integral is involved in the system, we will also denote the set of equations as integro-partial differential equation (IPDE) system.

We take a closer look at the involved parameter-functions:

μ,$\overline{\mu}$ Independent of whether a person is infected or not, there is always an age-dependent

chance that it dies. Hence,

$$\mu(a) : \mathbb{R}^+ \to \mathbb{R}^+ : \ a \mapsto \mu(a), \qquad \overline{\mu}(a) : \mathbb{R}^+ \to \mathbb{R}^+ : \ a \mapsto \overline{\mu}(a)$$

can be interpreted as a death-rate of an individual with age a if it is healthy (left) or sick (right). We usually assume

$$\overline{\mu}(a) \geq \mu(a)$$

holds, as sick individuals are usually more likely to die. Nevertheless, we will moreover focus on $\overline{\mu} \equiv \mu$ thinking of non-lethal diseases.

γ As the age is a direct indicator for the physical fitness of a person the recovery-chance γ depends on it. Therefore

$$\gamma : \mathbb{R}^+ \to \mathbb{R}^+ : \ a \mapsto \gamma(a)$$

can be interpreted as a recovery-rate of an individual with age a.

κ It was already mentioned that contacts between individuals sharing the same age are much more likely. Hence

$$\kappa : \mathbb{R}^+ \to \mathbb{R}^+ : (a_1, a_2) \mapsto \kappa(a_1, a_2)$$

can be interpreted as a rate that an individual with age a_1 gets in contact with an individual with age a_2.

Typical shapes for μ, κ and γ can be found applying the ideas right at the beginning of this section. For our theoretical analysis we used the fictional parameter functions defined in Figures 5.27 and 5.26.

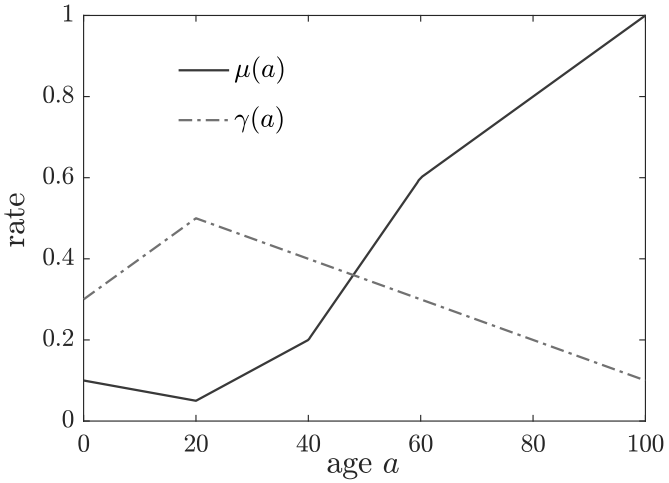


Figure 5.26: Fictional mortality-rate $\mu(a) \equiv \overline{\mu}(a)$ and recovery rate $\gamma(a)$.

We moreover state a related agent-based model.

Model 5.4: Agent-Based Age-Dependent SIR Model

Based on the mentioned parameter functions, we define an agent-based model via its initial setup and its time evolution.

Initial-Setup:

- The model consists of $N \gg 1$ agents $I_i, i \in \{1, \dots, N\}$.
- Each agent has a certain state $I_i(t)$ consisting of a health-state $(I_i(t))_1 \in \{S, I, R\}$ (i.e. *susceptible*, *infected* and *recovered*) and a certain age $(I_i(t))_2 \in \mathbb{R}^+$. The states are initially generated according to three density functions: First, a *health-state* is assigned according to the probabilities:

$$P((I_i(0))_1 = S) = \int_0^\infty \frac{S_0(a)}{N} da,\ P((I_i(0))_1 = I) = \int_0^\infty \frac{I_0(a)}{N} da,$$

and $(I_i(0))_1 = R$, otherwise. Afterwards their initial age $(I_i(0))_2$ is sampled according to the corresponding one of the three density functions

$$\frac{S_0(a)}{\int_0^\infty S_0(a)da}, \frac{I_0(a)}{\int_0^\infty I_0(a)da}, \frac{R_0(a)}{\int_0^\infty R_0(a)da}.$$

The model is simulated in equidistant time-steps with length dt starting at $t = t_0$ and stopping at t_{end}. Hereby the step-length dt needs to be small enough to guarantee

$$\sup_{a,b} \left(\max(dt \cdot \gamma(a), \quad dt \cdot \mu(a), \quad dt \cdot \overline{\mu}, \quad dt \cdot \kappa(a,b))\right) < 1$$

with the already defined parameter functions. Hence, $dt \cdot \gamma$, $dt \cdot \mu$, $dt \cdot \overline{\mu}$ and $dt \cdot \kappa$ can be used as probability-functions.

Dynamics:

- Each time-step each agent is addressed once.
- In case the agent's health-state is *susceptible* $(I_i(t))_1 = S$, a second, randomly picked agent I_j is investigated. With an age-dependent probability

$$\kappa((I_i(t))_2, I_2^j(t)) \cdot dt,$$

  agent i has contact with agent j. In case of a contact, if agent j was infected, agent i becomes infected too.
- In case the agent's health-state is *infected* $(I_i(t))_1 = I$, the agent recovers with probability $\gamma(I_2^i(t)) \cdot dt$.
- Furthermore, if the agent is infected, it dies with probability $\overline{\mu}((I_i(t))_2) \cdot dt$ otherwise with probability $\mu((I_i(t))_2) \cdot dt$. Hereby it leaves the model.
- Hereafter, all agents increase their age by dt.

The McKendrick equation can be proven to be a valid mean-field model for the stated microscopic approach. This mean-field analysis link has already been published in [Bicher et al., 2017a] and the reader is referred to this publication for getting deeper insights into this process as it is time-consuming and complex.

The similarity of the model results can be verified in Figure 5.28. For further contentual interpretation of the results the reader is once again referred to [Bicher et al., 2017a].

The case study in this section shows, that it is possible to find mean-field models for microscopic models with hybrid state-spaces and interaction between sub-models. Nevertheless, mean-field analysis processes for such models always turn out to be quite time consuming and complicated, especially for inexperienced modellers.

5.5.3 Interpretation of Aggregated Numbers in Space Continuous Models

In this section we want to focus on specific features of the aggregated numbers of state continuous microscopic models in general, that especially become interesting when thinking of a macroscopic mean-field model. Surprisingly, there are a couple of questions that arise when results of microscopic models with continuous state-space need to be interpreted on the aggregated level.

We study the microscopic model 5.4 and investigate the question:

$$\textit{How many agents are in state } (S, 10.5)\textit{?} \tag{5.101}$$

In reality, the corresponding question

$$\textit{How many susceptible persons are precisely } 10.5 \textit{ years old?} \tag{5.102}$$

will almost surely be answered with zero as age is measured on a continuous scale and the probability that an individual is precisely 10.5 years old at a given time-instant vanishes. In

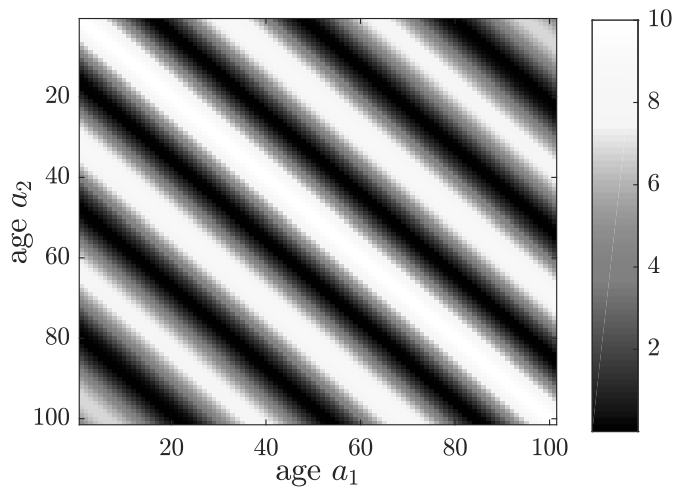


Figure 5.27: Heat-map of the fictional contact-rate $\kappa(a_1, a_2)$. The intended band-matrix structure appears because of parent-children-relationships.

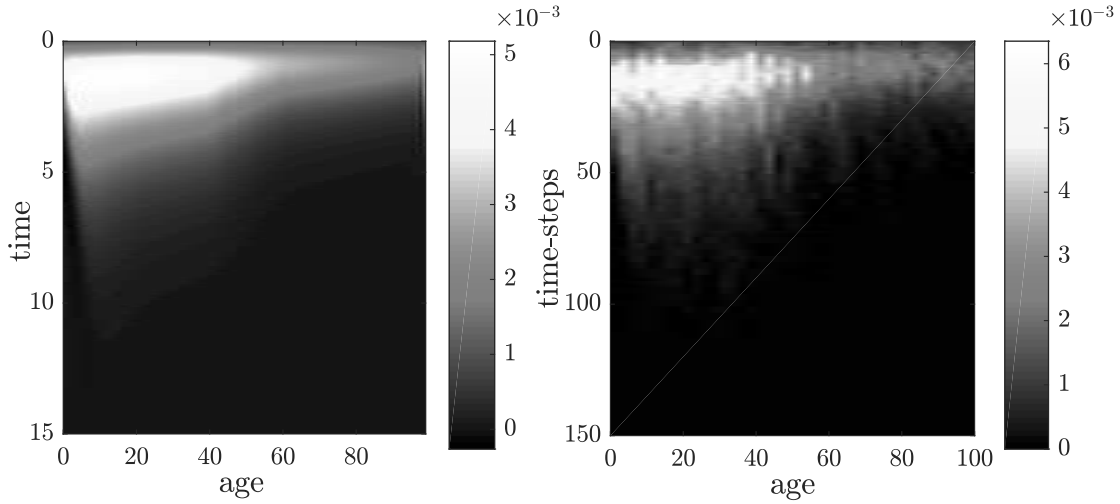


Figure 5.28: Results for the infected density $I(t)/N$ compared for both models with $N = 8000, dt = 0.1$ and a given initial condition. The parameters functions were used as defined in Figures 5.27 and 5.26. The ABM results can be seen to the right and are plotted as histogram with 40 bins. IPDE results gained with (modified) method of lines are shown to the left.

the presented age-dependent SIR model 5.4 (and in state-continuous microscopic models in general), we might observe the same behaviour, but not necessarily. We define three small model scenarios for Model 5.4 to clarify this:

a) Set
$$S_0(a) := \mathbb{1}_{[0,100]}(a)\frac{N}{100},\ I_0(a) :\equiv 0,\ R_0(a) :\equiv 0$$
and $dt := 0.5$. As $I_0(a) = R_0(a) = 0$, all agents are initially susceptible and will remain in this state, as no disease is present. Their initial age is sampled according to $S_0(a)/N$ which corresponds to the density function of the uniform distribution on the interval $[0, 100]$. As the uniform distribution is a continuous distribution, the sampled age of the agents $(I_i(0))_2$ will also fulfil
$$P((I_i(0))_2 = 10.5) = 0.$$
As the model is updated in steps with length dt and age is enhanced by this amount, also
$$P((I_i(k \cdot dt))_2 = 10.5) = 0, \quad k \in \mathbb{N}$$
holds. Consequently, the answer to above question is equivalent to the answer in reality.

b) Set $S_0(a) := N\delta(a)$, with the delta distribution δ, $I_0(a) :\equiv 0,\ R_0(a) :\equiv 0$ and $dt := 0.5$. Consequently, again, all agents start in the susceptible health state, but this time with age 0 (compare with the Dirac (probability) measure 1.4). Clearly, $P((I_i(0))_2 = 10.5) = 0$, but after 21 model steps with length 0.5 all agents are 10.5 years old and, therefore,
$$P((I_i(21 \cdot dt))_2 = 10.5) = 1.$$
Suddenly, the question about how many agents are in state $(S, 10.5)$ is not trivial anymore. Hence, the answer to this question differs from reality.

If one tries to simulate the macroscopic IPDE approach it will quickly turn out that the use of a delta distribution as initial condition is not the best choice to achieve a realistic model. Yet, this is not so obvious for microscopic models. When applying cohort studies, for example, one might easily be tempted to initialise a model with a considerably large number of agents with equivalent age to see how the birth cohort enhances with time.

In the random-walk model 5.1 the initial Dirac pulse (all particles started in $(0, 0)$) was a realistic initial condition as it nicely models a punctual source of pollution in a groundwater stream. The diffusion parameter, which is missing in the SIR model, additionally justified this choice.

c) Even if the modeller is aware that irregularities may occur if initial conditions are not chosen properly also other processes can lead to problems. Set $S_0(a) := \mathbb{1}_{[0,100]}(a)\frac{N}{100}$, $I_0(a) := 0,\ R_0(a) := 0$ and introduce a birth-process according to the microscopic Levins model 1.1 in Section 5.1.3. I.e. each time-step each agent has an age-dependent probability to create an offspring. Logically, the offspring is initialised susceptible with age zero.

At least for 20 time-steps $P((I_i(k \cdot dt))_2 = 10.5) = 0$ will be true, but at the 21$^{\text{st}}$, all agents that were born in the first step are precisely 10.5. Hence,

$$P((I_i(21 \cdot dt))_2 = 10.5) > 0.$$

Though the model is initialised properly, irregularities will be brought in the model if newly generated agents are not initialised correctly. One way to solve this problem is given as follows: Newborn agents in the k-th model step are assigned a uniformly sampled age between $[0, dt)$ indicating that they have been born at some time during the last time-step. We find this strategy later in the case study in Section 5.6.

Summarising, continuous variables in microscopic models need to be treated analogous to their real pendants. In reality, the correct question would clearly be:

How many susceptible persons have had their 10th birthday, but not yet their 11th? (5.103)

Consequently, also in the model we need to ask ourselves:

How many agents' states lie inside the set $\{(S, x), x \in [10, 11)\}$*?* (5.104)

Hereby the choice of the interval parentheses (open, closed) should not play a role if the model is defined properly (compare with the three scenarios above).

At this point, surprisingly a second difficulty arises, which seems obvious at first, but becomes particularly tricky at second glance: Is question (5.104) for the microscopic model the optimal pendant to question (5.103)?

The reason why this is not trivial is, that a model agent does not represent one specific real person, but acts as a **statistical representative** of a real person. That means, a snapshot of the model at a given time-instant, i.e. the vector of agents represented by their state, say

$$\vec{X} := \{X_1, X_2, \ldots, X_N\} := \left\{ \begin{pmatrix} S \\ 16.3 \end{pmatrix}, \begin{pmatrix} S \\ 10.1 \end{pmatrix}, \begin{pmatrix} S \\ 65.4 \end{pmatrix}, \ldots, \begin{pmatrix} S \\ 5.3 \end{pmatrix} \right\},$$

is not equivalent to a sample of real persons, but is at most **equally distributed**. Thus, in order to answer question (5.103) using the model, we have to find – i.e. estimate – the distribution of the X_i and, precisely, the distribution of $\sum_{i=1}^{N} X_i \sim NX_i$ and use the information about the estimated distribution to answer the question. Mathematically spoken this corresponds to (there might be other options as well), first, estimation of the density curve of the continuous distribution and , second, perform integration afterwards to get the statistically correct aggregated numbers.

Estimate a density function f *for the iid sample* X_i*. How large is the integral* $N \int_{10}^{11} f(x)dx$*?* (5.105)

As mentioned in Section 3.2.2 there are in principle two different strategies for this process.

The first one is the standard histogram. That means, based on specific bins, i.e. a set of disjoint, halve-open intervals that cover the observed space, the total number of agents in specific intervals are counted, sometimes normed, and depicted as a piecewise constant function

that represents an estimated probability density. **If** the bins of the histogram are chosen as the intervals $[x, x+1), x \in \mathbb{N}_0$, then this density-estimation and integration strategy is equivalent to answering question (5.104), **but** it is not wise to choose the set of bins for the histogram according to the given research question! The bins should rather be chosen to ensure the optimal fit of the estimated density, e.g according to [Scott, 1979].

Say, model 5.4 is run with $N = 100$ agents then a histogram with bins $[x, x+1), x \in \mathbb{N}_0$ is completely useless: We expect that age of persons are distributed on the interval $[0, 105)$ which would result in a histogram with 105 bins. Thus, on average, each histogram bin will contain less than one person which is completely unrepresentative. To estimate a representative density function, we have two options

1. rerun the model several times to enlarge the sample size, or
2. use larger bins, e.g. $[10x, 10x+10), x \in \mathbb{N}_0$.

The first of these two options is clearly preferred, but requires additional time for simulation-runs.

The second option to estimate density functions mentioned in Section 3.2.2 is the strategy of kernel density estimation. If a good bandwidth is chosen, e.g. via (3.25), kernel density estimation is the better choice to approximate a continuous density function for its smoothness. Moreover, it is possible to include additional information about the distribution to the density-estimation by choosing smaller or larger bandwidths or by using different kernel functions.

The macroscopic IPDE approach Model 5.3 can be used as a reference for the fit of the density estimation as it directly models the temporal development of the aggregated density curve. If the chosen density estimation of the microscopic model results (qualitatively) matches the result of the IPDE then the chosen bandwidth or the chosen number of bins was good. Figure 5.29 gives an example how estimated densities of microscopic model results may look like. The upper left plot shows the (numeric) results of the macroscopic model. The plot directly to its right shows a Gaussian kernel density estimation of the corresponding microscopic model with 1000 agents. The bottom row shows two histogram representations of the same result. Easily seen, the histogram on the left with 20 bins is qualitatively similar to the macroscopic density results, while the analogous histogram with 200 bins right next to it is completely unrepresentative.

5.6 Mean-Field Analysis for Parametrised Models

Throughout literature in social sciences mean-field analysis can be seen to be a rather theoretical tool for un-parametrised academical microscopic models – we usually find these kind of models pejoratively denoted as "toy-models" to indicate that they are hardly usable to deduce information for the real-world system, at least with respect to quantitative information. Yet, these models are vitally important for development, testing and benchmarking of methods. Examples for that are found in almost any field of application from sociology to medicine and especially epidemiology [Webb et al., 2007, Fibich and Gibori, 2010, de Aguiar et al., 2003, Lachapelle and Wolfram, 2011, Benoit et al., 2006, Edwards et al., 2003, Deffuant et al., 2000], but also the mod-

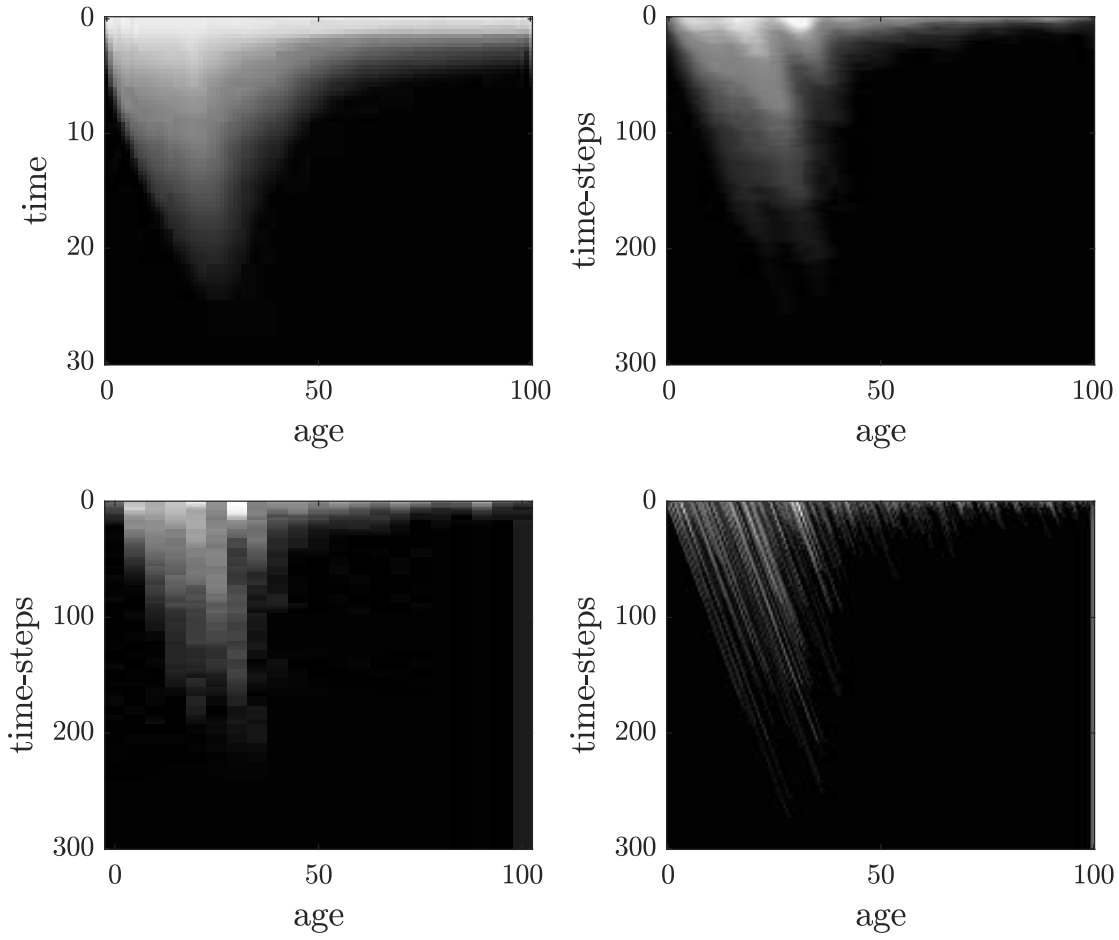


Figure 5.29: Visualisation of model results for the susceptible population of the age-dependent SIR model. Results of the IPDE model (Model 5.3) are seen at the top-left. Different density estimations of the microscopic model (Model 5.4) results with $N = 1000$ agents can be seen on the other three plots.

els stated in this work so far can in principle be counted to this class. Consequently, features like the correct choice of parameters, verification and validation of the models are not considered.

We assume that there are mainly two reasons for that: First, it seems as if experts who work on mean-field analysis are usually more interested in the method and the involved model comparison on the theoretical level, than on the application of the model as a real-life decision-support tool. The gap between high level, general methodological research on modelling and model development for decision-support applications is larger than ever due to specialisation of research groups. Secondly, it can be assumed that the time required for mean-field analysis of parametrised models, which are usually much more complex than mentioned academic models, would be too extensive. It is a fact that heterogeneity of the state-space (one may think of a small-world network instead of a rectangular grid) or more complex individual behaviour of sub-models (memory, goal oriented behaviour, adaptive behaviour) makes mean-field analysis extremely difficult as a lot of abstraction and cutback processes need to be made.

In this section we want to give an example for a mean-field analysis of a fully parametrised and validated population model. Hereby, we want to prove that the additional effort of performing mean-field analysis for parametrised models is reasonably large and that it can be a valuable analysis tool: We will show how the derived mean-field model helps with the parameter identification process of the model.

5.6.1 Case Study: Generic Population Concept (GEPOC)

The Generic POpulation COncept (GEPOC) is one of the most successful achievements of Comet research project DEXHELPP. The main goal of it was to develop a solid foundation of any (simulation) model that is somehow based on Austria's population. This includes, for example, models for epidemiological analysis, analysis of supply and demand in medical care, resource planning or health technology assessment. It is not simple to explain what physical form mentioned GEPOC really has. It can be seen to at least consist of

- an agent-based simulation model that depict the current, and forecasts the (close) future population of Austria,
- an analogous system dynamics model for the same purpose,
- a set of processed data sheets containing sufficient information for successful parametrisation of any similar population model for Austria, and finally
- a large (unpublished) *Generic Population Handbook* [Miksch et al., 2015] that contains any information about the first three points including a brief description of the models, the data acquisition and pre-processing process, the model validation and verification as well as a brief cross model validation with prognoses gained from the Austrian Bureau of Statistics, our main data-source in this project.

Shortly after the development of GEPOC was finished and the models were successfully validated, the agent-based model, henceforth referred to as **GEPOC ABM**, turned out to be the most applicable part of the listed GEPOC project outcomes. Due to its flexibility with respect

to model extensions it has already proven its worth as a base-model for several applications. We would like to mention the analysis of measles and polio vaccination rates in Austria (a publication is to appear soon) and the investigation of hospitalisation and re-hospitalisation of mental patients to psychiatric hospitals (to appear shortly in [Zauner et al., 2017]). Slightly more about the second application will be presented in the next section.

Anyway, for more specific details about GEPOC the reader is referred to [Bicher et al., 2015, Miksch et al., 2016] or to the Generic Population Handbook [Miksch et al., 2015], which may be obtained on request. Here we will state a definition of the agent-based model.

Model 6.1: Agent-Based GEPOC Model

As done several times before in this work, we define the model via its initial set-up and its dynamics.

Initial Setup:

- The model initially consists of N_0 agents $I^i, i = 1 \dots N_0$ of which $\overline{F_0}$ have the property *female* and $\overline{M_0}$ have the property *male*. We speak of male and female agents. According to given distributions, male and female agents are given an initial *age*. Hereby let $f_0(a)$ describe the density for the female agents while $m_0(a)$ describes the density for the male ones. As we will need it later on, we combine f_0 and $\overline{F_0}$ as well as m_0 and $\overline{M_0}$ to an aggregated density mapping via

$$F_0(a) := \overline{F_0} f_0(a), \quad M_0(a) := \overline{M_0} m_0(a). \tag{5.106}$$

  We write

$$I_i(t) = (s, a)$$

  indicating that agent I_i has sex s and age a at time-instant t.

The model is furthermore updated in equidistant time-steps of length dt [years]. Each time-step, the following actions take place. All agents are updated simultaneously. **Dynamics:**

- Each time step consists of two main parts.
- The first part each agent is addressed once. For each of them, the following actions are applied.
  - An initially empty list of actions is assigned, that happened to the agent in the course of the past time-step. This list is iteratively filled with actions:
  - There is a certain time, sex and age dependent probability

$$Dp_{dt}(t, s, a) \tag{5.107}$$

    that the agent with sex s and age a died in the past time-step – i.e. a "death-action" took place. Hence, a randomly sampled *death-date* is added to mentioned action list.

  - There is a time, sex and age dependent probability

$$Em_{dt}(t, s, a) \tag{5.108}$$

    that the agent with sex s and age a emigrated during the last time step. In that case, a randomly sampled *emigration-date* is added to the action list.

  - There is a time and age dependent probability

$$Bp_{dt}(t, a) \tag{5.109}$$

    that a female agent with age a generates an offspring. In that case, a randomly sampled *birth-date* is added to the action list.

  - After that, all actions are applied in the correct order. In case the agent dies or emigrates, it is removed and all future scheduled actions are cansiteed. For the birth-action, one new agent with a uniformly sampled age between 0 and dt is generated and added to the model at the end of the time-step. It is male with constant probability $Bf(m)$, female otherwise ($Bf(f) := 1 - Bf(m)$).

- The second part of the time-step consists of immigration. Hereby a fixed number of "immigrant-agents" are added to the model according to a time, age and sex dependent aggregated number

$$Im_{dt}(t, a, s). \tag{5.110}$$

  That means $\int_0^\infty Im_{dt}(t, a, f)da + \int_0^\infty Im_{dt}(t, a, m)da =: \overline{Im}^f_{dt}(t) + \overline{Im}^m_{dt}(t) = \overline{Im}_{dt}(t)$ new agents will be generated. First their sex is randomly chosen according to the fraction of $\overline{Im}^m_{dt}(t)$ and $\overline{Im}^f_{dt}(t)$, then (for sex s) their age is sampled according to the density

$$\frac{1}{\overline{Im}^s_{dt}(t)} Im_{dt}(t, a, s).$$

The strategy to use an event-based action-list for correct scheduling of (probably multiple) events that occur in the time-step perfectly maps the real causal logic and avoids the problem, in which order agent actions should be applied within the time-step. This problem is not trivial as in reality e.g. individuals might die after they gave birth to a child in a time interval, but cannot give birth to a child after they have died. In case one tries to define a fixed order for these two events, the probabilities will be distorted as soon as they occur both in one time-step.

In case of the population model this concept does not really pay off, as the probabilities that any two of the actions occur in one time-step are extremely small. Nevertheless, it is aim of the GEPOC ABM to pose for a flexible framework that can be extended by any kind of modules. The mentioned unlikeliness of two "simultaneous" events must not be given anymore – e.g. the risk of developing cancer and the risk of death are even positively correlated.

It is questionable if the regarded model should really be denoted as agent-based model as the most important feature of agents is missing: interaction. We, yet, deemed this terminology optimal, as the model is only a basic module for more advanced and larger models which, then,

may include interaction between the individuals.

As stated in [Bicher et al., 2015, Miksch et al., 2016] the model, usually executed with about 8 million agents (1 agent $\sim$ 1 person), is fully parametrised, verified and validated using data and prognoses (prognosis tool *Sikurs* [Austria, 2012]) gained from the Austrian Bureau of Statistics. Find details in [Bicher et al., 2015] and [Miksch et al., 2015]. Summarising, the model is able to validly depict Austria's population from 1990 to 2016 and generates credible prognoses up to several decades into the future. The liability of long term prognosis may be doubted (as well as any other demographic long term prognosis), nevertheless these population prognoses are the best base for planning we currently have.

5.6.2 Mean-Field Analysis of GEPOC ABM

Though the model is fully parametrised with real data, the basic model structure of GEPOC ABM is quite simple. Hence, it is legit to assume that there might be a valid mean-field model for it.

Surprisingly, the mean-field analysis can be done very similar to the already mentioned age-dependent SIR model in Section 5.5. We perform the analysis analogous to the step-by-step process described in Section 4.6.2, but do not specifically refer to each step.

1. First, we easily find $N(0) = N_0$ agents I_i at the beginning of the simulation. Their common state-space is investigated and fixed with

$$\Gamma := \{m, f\} \times \mathbb{R}^+. \tag{5.111}$$

   I.e. the state-space is a hybrid one. Thus, each agent has a time dependent state via

$$I_i(t) = (s, a),\ s \in \{m, f\},\ a \in \mathbb{R}.$$

2. We identify all possible transitions between agent states which results in Table 5.1

3. In the list only one real transition is visible, namely the ageing process. As each agent definitely ages by dt after each time-step in which it did not die or emigrate we get, for $b, c \in \mathbb{R}^+$

$$\begin{aligned} P(I(t+h) \in \{(s,x), x \in [b,c]\} | I(t) = (s,a)) \\ = \mathbb{1}_{[b,c]}(a+dt)(1 - Dp_{dt}(t,a,s))(1 - Ep_{dt}(t,a,s)) \\ \approx \mathbb{1}_{[b,c]}(a+dt)(1 - Dp_{dt}(t,a,s) - Ep_{dt}(t,a,s)). \end{aligned}$$

   The correspondent density is a Dirac pulse

$$\begin{aligned} \tilde{P}(I(t+h) = (s,x) | I(t) = (s,a)) \\ \approx \delta(x - a - dt)(1 - Dp_{dt}(t,a,s) - Ep_{dt}(t,a,s)). \end{aligned}$$

   Consequently, we define the transition kernel

$$\omega(t, (s,a), (s,x)) := \frac{\delta(x - a - dt)}{dt}(1 - Dp_{dt}(t,a,s) - Ep_{dt}(t,a,s)). \tag{5.112}$$

State Transition	Reason
$(m, a) \to \emptyset$	a male agent with age a dies
$(f, a) \to \emptyset$	a female agent with age a dies
$(m, a) \to \emptyset$	a male agent with age a emigrates
$(f, a) \to \emptyset$	a female agent with age a emigrates
$\emptyset \to (m, a + dt)$	a male agent with age a immigrates
$\emptyset \to (m, a + dt)$	a female agent with age a immigrates
$\emptyset \to (m, a)$	a male agent with age a ($a \in [0, dt)$) is born
$\emptyset \to (f, a)$	a female agent with age a ($a \in [0, dt)$) is born
$(m, a) \to (m, a + dt)$	a male agent ages
$(f, a) \to (f, a + dt)$	a female agent ages

Table 5.1: Possible state transitions of agents in GEPOC ABM Model 6.1. The symbol $\emptyset$ denotes that the agent either leaves the model or enters it.

4. We now consider birth processes in the model. Hereby we find one process wherein agents are offspring by other agents. As a newborn agent's age is uniformly sampled in $[0, dt)$ we receive the density

$$\tilde{P}(\text{new agent born in state } (s, x) | I(t) = (f, a)) = \frac{\mathbb{1}_{[0,dt)}(x)}{dt} Bp_{dt}(t, a) Bf(s)$$

and therefore

$$c(t, (s, x), (f, a)) := \frac{\mathbb{1}_{[0,dt)}(x)}{dt^2} Bp_{dt}(t, a) Bf(s). \tag{5.113}$$

5. A second, global, creation rate was found: immigration. In this model, a time-step is too large to ask for the probability that one agent enters the model (which is almost surely one) but we include multiple agents at once. Hence, we need to apply a trick as the mean-field theorem is not applicable for this case:

   Define a sufficiently large $Q \in \mathbb{N}$ and rephrase the immigration process of the microscopic model a little bit. Instead of only regarding one immigration process we regard Q different immigration processes of which each may only lead to the immigration of one person with probability

$$\frac{\int_0^\infty Im_{dt}(t, a, f) + Im_{dt}(t, a, m) da}{Q}.$$

   The densities by which the immigrant's age and sex are decided remain the same (compare with (5.110)). Clearly, the rephrased model behaves like the original on the average and the added fluctuations vanish for large numbers of immigrants ($\overline{Im}_{dt}$) by the Law of Large Numbers.

   Hence, for each of those processes $j \in \{1, \ldots, Q\}$ we calculate the immigration density

$$\tilde{P}(\text{new agent immigrated in state } (s, x) \text{ in process } j) = \frac{Im_{dt}(t, x, s)}{Q}$$

leading to the creation kernel

$$C(t,(s,a)) := \sum_{j=1}^{Q} \frac{Im_{dt}(t,x,s)}{Q \cdot dt} = \frac{Im_{dt}(t,x,s)}{dt}, \tag{5.114}$$

by summing all creation processes.

6. Finally, death as well as emigration equally leads to removal of single agents. Hence, considering that agents only die if they did not emigrate and vice versa,

$$\begin{aligned} &P(\text{death or emigration}|I(t)=(s,a)) \\ &\quad = Dp_{dt}(t,a,s)(1-Ep_{dt}(t,a,s)) + Ep_{dt}(t,a,s)(1-Dp_{dt}(t,a,s)) \\ &\quad \approx Dp_{dt}(t,a,s) + Ep_{dt}(t,a,s). \end{aligned}$$

Finally, we set

$$d(t,(s,a)) := \frac{Dp_{dt}(t,a,s) + Ep_{dt}(t,a,s)}{dt}. \tag{5.115}$$

7. Putting equations (5.112),(5.113),(5.114) and (5.115) together as it is defined in the step-by-step process, we obtain an IPDE system. Writing $\varphi_f(a,t)$ as the density of female agents and $\varphi_m(a,t)$ as the density for the male ones, we obtain

$$\begin{aligned} \frac{\partial}{\partial t}\varphi_f(a,t) = C_f - \varphi_f(a,t)d_f + \int_0^\infty \varphi_f(b,t)c_f(b)db \\ + \int_0^\infty \varphi_f(b,t)w_f(b,a) - \varphi_f(a,t)w_f(a,b)db \end{aligned} \tag{5.116}$$

$$\begin{aligned} \frac{\partial}{\partial t}\varphi_m(a,t) = C_m - \varphi_m(a,t)d_m + \int_0^\infty \varphi_f(b,t)c_m(b)db \\ + \int_0^\infty \varphi_m(b,t)w_m(b,a) - \varphi_m(a,t)w_m(a,b)db. \end{aligned} \tag{5.117}$$

with the abbreviations ($s \in \{m,f\}$)

$$C_s := C(t,(s,a)) = \frac{1}{dt} Im_{dt}(t,a,s), \tag{5.118}$$

$$c_s(b) := c(t,(s,b),(f,a)) = \frac{1}{dt}\frac{\mathbb{1}_{[0,dt)}(b)}{dt} Bp_{dt}(t,a)Bf(s), \tag{5.119}$$

$$d_s := d(t,(s,a)) = \frac{1}{dt}(Dp_{dt}(t,a,s) + Ep_{dt}(t,a,s)), \tag{5.120}$$

$$w_s(a,b) := \frac{\delta(b-a-dt)}{dt}(1 - Dp_{dt}(t,a,s) - Ep_{dt}(t,a,s)). \tag{5.121}$$

As mentioned it is a good idea to perform a Taylor approximation for $\varphi_f(b)$ and $\varphi_m(b)$ at a in the second integral terms as $\omega(a,b) = 0$ for large $|a-b|$. Hereby we are able to

pull these terms out of the integral. Note, that it is not wise to do the same for the first integral term as the age difference between a potential mother (age b) and her child (age a) is usually quite large.

Therefore, exemplary for φ_m, but analogously for φ_f,

$$\varphi_m(b,t) \approx \varphi_m(a,t) + \frac{\partial}{\partial a}\varphi_m(a,t)(b-a) + \frac{\partial^2}{\partial a^2}\varphi_m(a,t)\frac{(b-a)^2}{2}$$

can be used as approximation. We obtain

$$\begin{aligned}
&\int_0^\infty \varphi_m(b,t)w_m(b,a) - \varphi_m(a,t)w_m(a,b)db \\
&\quad\approx \varphi_m(a,t)\underbrace{\int_0^\infty w_m(b,a) - w_m(a,b)db}_{=:I} + \frac{\partial}{\partial a}\varphi_m(a,t)\underbrace{\int_0^\infty (b-a)w_m(b,a)db}_{=:II} \\
&\qquad\qquad + \frac{\partial^2}{\partial a^2}\varphi_m(a,t)\underbrace{\int_0^\infty \frac{(b-a)^2}{2}w_m(b,a)db}_{=:III}.
\end{aligned}$$

Dealing with the single parts we get (now, again, for $s \in \{f,m\}$)

$$\begin{aligned}
I &= \int_0^\infty \frac{\delta(a-b-dt)}{dt}(1 - Dp_{dt}(t,b,s) - Ep_{dt}(t,b,s)) \\
&\qquad - \frac{\delta(b-a-dt)}{dt}(1 - Dp_{dt}(t,a,s) - Ep_{dt}(t,a,s))db \\
&= \frac{1}{dt}\left(Dp_{dt}(t,a,s) - Dp_{dt}(t,a-dt,s) + Ep_{dt}(t,a,s) - Ep_{dt}(t,a-dt,s)\right) \\
&\qquad\qquad \approx \frac{d}{da}Dp_{dt}(t,a,s) + \frac{d}{da}Ep_{dt}(t,a,s), \quad (5.122)
\end{aligned}$$

$$\begin{aligned}
II &= \int_0^\infty (b-a)\frac{\delta(a-b-dt)}{dt}(1 - Dp_{dt}(t,b,s) - Ep_{dt}(t,b,s))db \\
&= -\frac{dt}{dt}(1-Dp_{dt}(t,a-dt,s)-Ep_{dt}(t,a-dt,s)) \approx -(1-Dp_{dt}(t,a,s)-Ep_{dt}(t,a,s)),
\end{aligned} \quad (5.123)$$

and

$$\begin{aligned}
III &= \int_0^\infty \frac{(b-a)^2}{2}\frac{\delta(a-b-dt)}{dt}(1 - Dp_{dt}(t,b,s) - Ep_{dt}(t,b,s))db \\
&= \frac{dt^2}{2dt}(1-Dp_{dt}(t,a-dt,s)-Ep_{dt}(t,a-dt,s)) \approx \frac{dt}{2}(1-Dp_{dt}(t,a,s)-Ep_{dt}(t,a,s)).
\end{aligned} \quad (5.124)$$

As $III = \mathcal{O}(dt)$ we will ignore it henceforth.

Using the Taylor approximation we finally obtain the following IPDE system.

Model 6.2: GEPOC Mean-Field Model

The mean-field model for GEPOC ABM is defined by the integro-partial differential equations

$$\frac{\partial}{\partial t}\varphi_f(a,t) = C_f - \varphi_f(a,t)(d_f - \alpha_f) + \int_0^\infty \varphi_f(b,t)c_f(b)db - \frac{\partial}{\partial a}\varphi_f(a,t)\beta_f \tag{5.125}$$

$$\frac{\partial}{\partial t}\varphi_m(a,t) = C_m - \varphi_m(a,t)(d_m - \alpha_m) + \int_0^\infty \varphi_f(b,t)c_m(b)db - \frac{\partial}{\partial a}\varphi_m(a,t)\beta_m \tag{5.126}$$

with the abbreviations ($s \in \{m, f\}$)

$$C_s := C(t,(s,a)) = \frac{1}{dt} Im_{dt}(t,a,s), \tag{5.127}$$

$$c_s(b) := c(t,(s,b),(f,a)) = \frac{1}{dt}\frac{\mathbb{1}_{[0,dt)}(b)}{dt} Bp_{dt}(t,a)Bf(s), \tag{5.128}$$

$$d_s := d(t,(s,a)) = \frac{1}{dt}(Dp_{dt}(t,a,s) + Ep_{dt}(t,a,s)), \tag{5.129}$$

$$\alpha_s := \frac{d}{da} Dp_{dt}(t,a,s) + \frac{d}{da} Ep_{dt}(t,a,s) \tag{5.130}$$

$$\beta_s := 1 - Dp_{dt}(t,a,s) - Ep_{dt}(t,a,s). \tag{5.131}$$

The initial values and boundary conditions are given by

$$\varphi_f(a,0) = F_0(a), \quad \varphi_m(a,0) = M_0(a), \quad \frac{\partial}{\partial a}\varphi_f(0,t) = \frac{\partial}{\partial a}\varphi_m(0,t) = 0, \tag{5.132}$$

as the density needs to be conserved.

For small dt the probabilities Dp_{dt} and Ep_{dt} are obviously small. Hence, $\alpha_s \approx 0$ and $\beta_s \approx 1$ may be assumed, which was done in discussion paper [Bicher and Popper, 2016].

The stated IPDE system is a valid mean-field model, but numerically very unstable due to the steep unsteady indicator function in the birth kernel c_s. We used a modified one-dimensional Method of Lines (MoL) with a very fine spatial discretisation and the implicit 5$^{\text{th}}$ order Backwards Differential Formula for time-integration. The integral term was solved by numerical integration with trapezoidal rule.

Though the ABM is fully parametrised we could find a suitable man-field model for it. Finally, we benefited from the established differential equation model in two interesting ways:

First, the derived model offered us a new view on the results of the microscopic model. While we mainly investigated cross sections of the microscopic model result (e.g. the temporal development of the number of male agents with birth year 1990), the results of the macroscopic mean-field model broadened our focus onto the population as a time-evolving, three dimensional

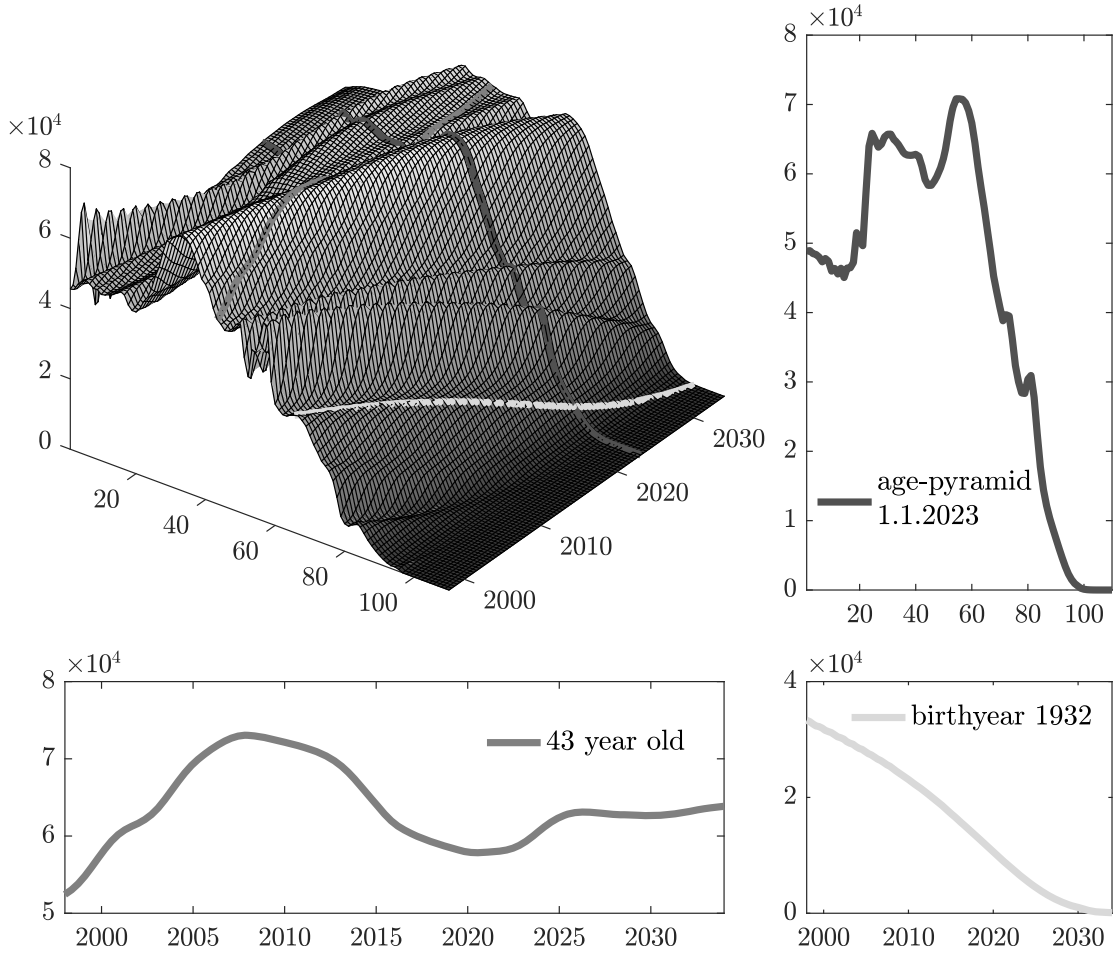


Figure 5.30: Visualisation of some numerically gained results of the GEPOC IPDE model 6.2. The density of the male population $M(a,t)$ is shown in the upper-left corner, different cross-sections are shown in the smaller plots around it.

density as seen in the upper-left plot of Figure 5.30. This view, though not usable for quantitative analysis, yet gives a nice picture of why Austria's population develops the way it does. One example for this is the impact of the so-called *baby boom* in the 1960 which leaves a visible peak in Figure 5.30 in the age pyramid. It can be seen in 1998 for the 40 year-old males and smoothly shifts the 60 year old in 2023.

Second, the derived model helped us with the calibration of a model extension. We will discuss this in more detail in the following section.

5.6.3 Calibration Case Study: Morbidity Extension of GEPOC ABM

A commonly known problem of large stochastic agent-based models like GEPOC ABM is that execution of the model is very time consuming. A single simulation run of the Python 3.6 implementation of GEPOC ABM with about 8 Million agents and 1500 time-steps, corresponding to either 30 years with $dt = 1$ week or about 4 years with $dt = 1$ day, requires about 30 minutes. Hereby it generates one stochastic simulation result which usually implies that the simulation needs to be repeated several times to get rid of the fluctuations. This issue has already been mentioned to be problematic when it comes to parameter identification of microscopic models and in specific, parameter **calibration**.

We explain the basics of a calibration problem: In case a parameter or parameter-vector value, say, the value of $\vec{p}$, is not precisely known from the beginning of the parametrisation process of a model, e.g. by surveys of measurements, it must be estimated with other means. Hereby it is a matter of the validation data if it is possible to find precise values for the parameter or if educated guesses from experts are the best you can get.

Say, for a specific simulation scenario A the modeller has access to real data, henceforth called **reference-data**, that would correspond to the desired (parts of the) simulation output of this scenario, then the modeller may adjust (i.e. **calibrate**) the parameter values of the model, so that the simulation output matches the given data. Moreover, the calibrated parameters can be used for different model scenarios $B, C, \ldots$ for which a reference output is not given. To formalise this in a more mathematical language we introduce the following notation. Suppose $S(\vec{p})$ corresponds to the simulated output of scenario A and Y denotes the desired simulation output according to the reference data, the calibration problem is formalised as

$$\vec{p}_{opt} = \text{argmin}_{\vec{p}\in X}(\|S(\vec{p}) - Y\|). \tag{5.133}$$

Hereby $\|\cdot\|$ denotes a suitable error measure. If this minimum is found, $\vec{p}_{opt}$ corresponds to the optimal parameter value for $\vec{p}$. As in case of reasonably complex simulation models impossible to calculate this minimum by hand, so-called **calibration algorithms**, usually iterative meta-heuristics, are used. I.e. one calculates a sequence $\vec{p}_i$ of parameter vectors that fulfil

$$\|M(\vec{p}_{i+1}) - Y\| \leq \|S(\vec{p}_i) - Y\|,$$

at least for large i, so that the sequence converges towards the optimal value.

Simulated Annealing, Generic or Evolutionary Algorithm or Particle Swarm Optimisation are probably the three most commonly known members of the large family of calibration algorithms. More about these and other calibration methods can be found e.g. in [Rippinger et al., 2016,Rippinger, 2016]. Though they are highly optimised and well-studied meta-heuristics, they all share the issue that successful calibration of a parameter value requires iterative execution of the simulation thousands of times.

During a study for the investigation of psychiatric diseases the agent-based GEPOC model was used and extended to depict hospitalisation and re-hospitalization numbers of psychiatric patients in Austria, Slovenia and parts of Italy. Hereby re-hospitalisation addresses the readmission of a psychiatric patient to a hospital after a certain time-period. The prediction of the total number of re-hospitalizations per country until year 2040 was the objective of the model. Without going into technical details about the three-staged model extension, for which the reader is

referred to [Zauner et al., 2017], calibration of specific parameters was required in several parts of it.

Hereby two main problems were detected that made the calibration of the models particularly difficult:

1. As defined, GEPOC ABM is a stochastic model. Hence, we have a slightly different calibration focus here. Instead of (5.133), we need to consider

$$\vec{p}_{opt} = \text{argmin}_{\vec{p}\in X}(\|\mathbb{E}(S(\vec{p}) - Y)\|). \tag{5.134}$$

   In principle, each time, the simulation is called by a calibration algorithm to calculate the error to the reference, it needs to be repeated several times to eliminate the fluctuations.

2. Already without the extension to a model for psychiatric diseases, the simulation of GEPOC ABM is very time consuming. Hence, execution of thousands of simulation-runs is hardly affordable with respect to time and computational resources.

In summary, a different strategy needs to be used. We applied a strategy based on mean-field theory which will be published soon in [Bicher et al., 2017b]. As it would be too confusing to explain this strategy directly on the example of the high-level re-hospitalisation model, we will define a **surrogate problem**. The calibration task in this surrogate model will in principle match the calibration tasks in the re-hospitalisation models, but spares confusing technical details.

Based on GEPOC ABM, we aim to include a non-infectious background morbidity that has a negative impact on the mortality rate of an agent. Yet, in the course of the model extension, the total population should be conserved.

The defined scenario is totally academic, but not so unrealistic. One may think of non-transmittable diseases like e.g. diabetes or, rarer, Chron's disease. Alternatively, one may also think of other mortality risk factors like smoking or heavy drinking which, in this sense, act like non-transmittable background diseases as they also have a negative impact on the mortality risk. The corresponding model is defined as follows:

Model 6.3: Background Morbidity Extension of GEPOC ABM

We will define the model extension based on the definition of GEPOC ABM 6.1 and change and add specific parts in its definition.

- The model initially consists of N_0 agents $I^i, i = 1 \dots N_0$ of which $\overline{F_0^+}$ have the properties **female and healthy**, $\overline{F_0^-}$ have the properties **female and morbid**. Analogously $\overline{M_0^+}$ and $\overline{M_0^-}$ have the properties **male and healthy** or **male and morbid**, respectively. According to given distributions, male and female agents are given an initial age. Hereby let $f_0^+(a)$ describe the density for the female and healthy agents, f_0^-, m_0^+ and m_0^- describe the corresponding densities for the other states. We com-

bine, exemplary, f_0^+ and $\overline{F_0^+}$ to an aggregated density mapping via

$$F_0^+(a) := \overline{F_0^+} f_0^+(a),\ F_0^-(a) := \overline{F_0^-} f_0^-(a), \tag{5.135}$$

$$M_0^+(a) := \overline{M_0^+} m_0^+(a),\ M_0^-(a) := \overline{M_0^-} m_0^-(a). \tag{5.136}$$

- We write

  $$I_i(t) = (s, h, a)$$

  indicating that agent I_i has sex $s \in \{f, m\}$, healthy state $h \in \{+, -\}$ and age a at time-instant t.

The following actions have changed in the dynamics of the model.

- In principle, both, healthy and morbid agents are treated equivalently, with exception of the death process. Hereby infectious agents have a higher death probability

  $$\xi(-) Dp_{dt}(t, s, a), \quad \xi(-) > 1. \tag{5.137}$$

  We assume that the time-step length is large enough to guarantee that the overall death-probability remains smaller than 1. Consequently

  $$\xi(+) Dp_{dt}(t, s, a) \tag{5.138}$$

  describes the death probability for the healthy agents. Herein $\xi(+)$ is not yet known. Therefore, its estimation is the **aim of the calibration**.

- Every healthy agent has an additional probability

  $$Mp_{dt}(t, s, a) \tag{5.139}$$

  to become morbid in the regarded time interval. This process is treated equivalently to all other agent-related processes in the model (i.e. sampling of an *infection-date* etc.).

- Newborn agents are always assumed to be healthy.

- The aggregated density for immigrated agents analogously splits from $Im_{dt}(t, a, s)$ to $Im^+(t, a, s)$ and $Im^-(t, a, s)$. They sum up to the values of the basic model $Im^+ + Im^- = Im$.

The estimation of the value of the unknown parameter $\xi(+)$ is goal of a calibration study. As the mortality for morbid agents is higher than in the model without morbidity, we clearly expect $0 < \xi(+) < 1$ to conserve the overall "level or mortality" and thereby the total number of agents in the model. As mentioned, we will use mean-field analysis of the model to find a way to correctly determine the parameter's value to conserve the overall population.

It is very simple to extend the mean-field analysis of GEPOC ABM to the background

morbidity model. First of all, the discrete part of the state-space is increased from $\{f, m\}$ to $\{f^+, f^-, m^+, m^-\}$. Hence, we expect a differential equation model with four instead of two equations. All other parts of the new mean-field model are rather self-explanatory extensions to the original IPDE model 6.2.

Model 6.4: Background Morbidity Mean-Field Model

$$\begin{aligned}
\frac{\partial}{\partial t}\varphi_f^+(a,t) =& C_f^+ - \varphi_f^+(a,t)(d_f^+ - \alpha_f^+) - \varphi_f^+(a,t)\gamma_f \\
&+ \int_0^\infty (\varphi_f^+(b,t) + \varphi_f^-(b,t))c_f(b)db - \frac{\partial}{\partial a}\varphi_f^+(a,t)\beta_f^+ \\
\frac{\partial}{\partial t}\varphi_f^-(a,t) =& C_f^- - \varphi_f^-(a,t)(d_f^- - \alpha_f^-) + \varphi_f^+(a,t)\gamma_f \\
&- \frac{\partial}{\partial a}\varphi_f^-(a,t)\beta_f^- \\
\frac{\partial}{\partial t}\varphi_m^+(a,t) =& C_m^+ - \varphi_m^+(a,t)(d_m^+ - \alpha_m^+) - \varphi_m^+(a,t)\gamma_m \\
&+ \int_0^\infty (\varphi_f^+(b,t) + \varphi_f^-(b,t))c_m(b)db - \frac{\partial}{\partial a}\varphi_m^+(a,t)\beta_m^+ \\
\frac{\partial}{\partial t}\varphi_m^-(a,t) =& C_m^- - \varphi_m^-(a,t)(d_m^- - \alpha_m^-) + \varphi_m^+(a,t)\gamma_m \\
&- \frac{\partial}{\partial a}\varphi_m^+(a,t)\beta_m^-
\end{aligned}$$

with the abbreviations ($s \in \{m, f\}$,$h \in \{+, -\}$)

$$\begin{aligned}
C_s^h &:= C(t,(s,h,a)) = \frac{1}{dt} Im_{dt}^h(t,a,s), \\
c_s(b) &:= c(t,(s,b),(f,a)) = \frac{1}{dt}\frac{\mathbb{1}_{[0,dt)}(b)}{dt} Bp_{dt}(t,a)Bf(s), \\
d_s^h &:= d(t,(s,h,a)) = \frac{1}{dt}(\xi(h)Dp_{dt}(t,a,s) + Ep_{dt}(t,a,s)), \\
\alpha_s^h &:= \xi(h)\frac{d}{da}Dp_{dt}(t,a,s) + \frac{d}{da}Ep_{dt}(t,a,s) \\
\beta_s^h &:= 1 - \xi(h)Dp_{dt}(t,a,s) - Ep_{dt}(t,a,s) \\
\gamma_s &:= \frac{1}{dt}Mp(t,a,s).
\end{aligned}$$

The initial values and boundary conditions are given by ($h \in \{+, -\}, s \in \{f, m\}$)

$$\varphi_f^h(a,0) = F_0^h(a), \quad \varphi_m^h(a,0) = M_0^h(a), \quad \frac{\partial}{\partial a}\varphi_s^h(0,t) = 0. \tag{5.140}$$

In order to conserve the total sum of the population we need to compare the integral of the sums of all functions φ_f and φ_m in Model 6.2 with the integral of sums of all functions $\varphi_f^+, \varphi_f^-, \varphi_m^+, \varphi_m^-$ in the upper IPDE model and determined $\xi(+)$ so that the resulting values are equivalent.

Clearly, above model is too complicated for direct calculation of $\xi(+)$ via formula manipulation. The interesting part of the strategy we present here is that this is not even the goal:

First, we summarise

$$\vec{V}(a,t) := (\varphi_f^+(a,t), \varphi_f^-(a,t), \varphi_m^+(a,t), \varphi_m^-(a,t)).$$

Quick investigation of the IPDE system in Model 6.4 makes it clear that there are vector valued functions $\vec{\Theta}$ and $\vec{\Xi}$ so that the system can be written as

$$\frac{\partial}{\partial t}\vec{V}(a,t) = \vec{\Theta}\left(\vec{V}, \frac{\partial}{\partial a}\vec{V}, a, t\right) + \xi(+)\vec{\Xi}\left(\vec{V}, \frac{\partial}{\partial a}\vec{V}, a, t\right), \tag{5.141}$$

wherein $\xi(+)$ only occurs as the prefactor of the second term. Consequently, also the integral of the sum over all coordinates $S(t) := \int_a \sum_{i=1}^4 \vec{V}_i(a,t)da$ can be expressed as

$$\frac{\partial}{\partial t}S(t) = \Theta\left(\vec{V}, \frac{\partial}{\partial a}V, t\right) + \xi(+)\Xi\left(\vec{V}, \frac{\partial}{\partial a}V, t\right). \tag{5.142}$$

with two scalar functions Θ and ξ as the integral is a linear operator. Hence, the **impact** of $\xi(+)$ on the time differential of the total population **is a linear one**.

This information seems strange at a first glance, but is indeed very valuable considering a calibration process of the microscopic model. Let henceforth $S(t, \xi(+))$ stand for the total population of the Background Morbidity model at time t, and let $S(t)$ denote the reference population of the GEPOC ABM model. Above analysis of the macroscopic model revealed the existence of a linear relationship of the expected difference $\mathbb{E}(S(t+dt, \xi(+)) - S(t, \xi(+)))$ and parameter $\xi(+)$. I.e.

$$\mathbb{E}(S(t+dt, \xi(+)) - S(t, \xi(+))) = d(t) + k(t)\xi(+)$$

for some scalar unknown functions $d(t)$ and $k(t)$, which can be used to find

$$\begin{aligned}\mathbb{E}(S(t+dt, \xi(+)) - S(t, \xi(+))) = k(t)\xi(+) + d(t) &\stackrel{!}{=} \mathbb{E}(S(t+dt) - S(t)) \\ \Rightarrow \xi(+) &= \frac{\mathbb{E}(S(t+dt) - S(t)) - d(t)}{k(t)}.\end{aligned} \tag{5.143}$$

We can draw two important conclusions: First, there is an easy method to determine the correct values for $k(t)$ and $d(t)$ thinking about properties of scalar linear affine functions: A line is uniquely determined by two points! So simulating the value of $\mathbb{E}(S(t+dt, x_1) - S(t, x_1))$ and $\mathbb{E}(S(t+dt, x_2) - S(t, x_2))$ for any two (reasonable) values x_1 and x_2 for $\xi(+)$ makes it possible to calculate the parameters of the affine transformation with very basic means.

Second, either $\xi(+)$ needs to depend on time (i.e. $\xi(+) = \xi(+, t)$) as $k(t)$ and $d(t)$ are very unlikely to be equivalent for all t, or we aim for an approximation $\xi(+)$ that does not perfectly

conserve the population for all times. It can be estimated by an optimisation algorithm when iteratively applying the differences $k(t)\xi(+) + d(t)$ on the initial population.

For the background morbidity model e.g. $x_1 = 1$ and $x_2 = 0.5$ are reasonable choices to generate the two sampling-points that uniquely define the linear affine function as only $\xi(+) \in [0, 1)$ makes sense. Thus,

$$\begin{aligned}
\mathbb{E}(S(t+dt,1) - S(t,1)) =& k(t) + d(t) \\
\mathbb{E}(S(t+dt,0.5) - S(t,0.5)) =& 0.5k(t) + d(t) \\
& \Rightarrow \\
k(t) =& 2\left(\mathbb{E}(S(t+dt,1) - S(t,1)) - \mathbb{E}(S(t+dt,0.5) - S(t,0.5))\right) \\
d(t) =& -\mathbb{E}(S(t+dt,1) - S(t,1)) + 2\mathbb{E}(S(t+dt,0.5) - S(t,0.5)) \\
\xi(+,t) =& \frac{\mathbb{E}(S(t+dt) - S(t)) - d(t)}{k(t)}.
\end{aligned}$$

The calibration strategy can be summarised (and generalised) as follows:

1. Establish a mean-field model and find a relationship $r(\vec{p}(t))$, how the unknown parameter (-vector) $\vec{p}(t)$ influences the differentiated aggregated numbers of the model.

   In the regarded case, $r(\vec{p}(t))$ was a linear affine function.

2. Execute the simulation with that many different parameter values $(\vec{p}(t) := \vec{x}_i)_{i=1}^d$ so that the parameters of r can uniquely be determined by the equation system

   $$\mathbb{E}(S(t+dt, \vec{x}_i) - S(t, \vec{x}_i)) = r(\vec{x}_i), \quad i = 1 \dots d. \tag{5.144}$$

   Hereby multiple simulations runs each need to be used to approximate the right-hand side expected value.

3. Invert r and apply r^{-1} on your reference differential data $S(t + dt) - S(t)$ to get the correct value for $\vec{p}(t)$ by

   $$\vec{p}_{opt}(t) := r^{-1}(S(t + dt) - S(t)). \tag{5.145}$$

The presented strategy seems very strange, but works extremely well, especially if linear relationships between parameter and aggregated numbers are present. In the course of developing the re-hospitalisation model this strategy made it possible to find calibrated parameter values by estimating only two expected values via the sample mean of 100 simulation-runs each, instead of thousands of iterations with a calibration meta-heuristic. Moreover, the Law of the Iterated Logarithm (and in specific its direct consequence (3.36)) can be used to find the quality of the calibrated fit by estimating the convergence rate of the sample mean.

As mentioned, this study is to appear in slightly different form in early 2018 in [Bicher et al., 2017b]. Herein the strategy will be presented omitting the mean-field analysis part solely focussing on the involved probabilities. Yet, the concept in behind is the same: **stochastic analysis of agent-based models leads to simple relationships for the aggregated numbers**.

CHAPTER 6

Classification of Microscopic Models

In this chapter we will finally summarise all findings that have been made in the presented mean-field approximation-based studies in Chapter 5. We propose a possible classification of microscopic models with respect to their aggregated system behaviour. Hereby we want to enrich the terminology that is currently used to describe microscopic models by very specific terms that convey an image how the model behaves as a system.

We deem this to be an important contribution as the standard terminology for labelling microscopic models, that means primarily the approaches described in Section 2.1, is not always perceived unambiguously in different fields of research. For example we discussed the usage of the term microsimulation models in Section 2.1.4 and found completely different meanings of this term for different application fields. It is not even clear if specific methods are only special cases of the other: In Section 2.1.5 we mentioned that there are modellers (e.g. [Law, 2007]) that interpret agent-based models as a special case of microscopic discrete event models, while there are also arguments that the situation might also be actually vice versa – there is no unique truth. Also the distinction between cellular automata and agent-based modelling can be tricky: The Susceptible-Infectious-Recovered (SIR) model presented in Section 5.2.1 could be interpreted as a lattice-gas cellular automaton, but could also be seen as a special case of an agent-based model on a grid.

A second aspect that motivates to find a more technical classification of microscopic models is that the statement "we developed an agent-based-, cellular automaton-, microsimulation-,... model" hardly indicates what challenges need to be solved when working with the model. This includes the questions

- how sensitive the model is,
- how time is updated in the model,
- what kind of interaction among the sub-models is given, or
- how model output needs to be interpreted.

6.1 Classification Parameters

Based on rigorous theoretical research on mean-field analysis and a great number of case studies we detected **four parameters**, which have a significant influence on the model's behaviour and the challenges involved with modelling and simulation. All of them can be detected solely based on the formal definition of the model (probably using some of the stochastic methods described in Chapter 3) and do not require any empirical tests with the model via simulation (e.g. like a sensitivity analysis).

6.1.1 Classification with Respect to Time-Update

The way how time is updated in a microscopic simulation model is very important for anyone who is not familiar with the technical details of the model. As mentioned several times in the course of this thesis there are in principle two types of microscopic models with respect to time update.

Time-discrete microscopic models are models that are updated in discrete time-steps of predefined length. Herein we may distinguish between models that update in equidistant time steps and models that use adaptive lengths of time-steps. All models discussed in this work use equidistant time-step widths, yet the original parametrised agent-based GEneric-POpulation Concept (GEPOC) model presented in Section 5.6.1 can be updated validly with arbitrary time-step length which also includes the usage of different length of time steps in one simulation run [Bicher et al., 2015].

In general, these kind of models need to be parametrised with information that explains which processes took place in the course of the skipped time interval in between the time-steps. In most cases, all sub-models are addressed once per time-step, but also different concepts can be found: In [Deffuant et al., 2000] precisely two (randomly chosen) sub-models are addressed per time-step. Updates of the sub-models states are usually performed simultaneously to avoid that certain sub-models have different opportunities only because the are addressed earlier or later than others.

In terms of time-discrete models two interesting challenges were discussed in the course of this thesis:

- Rescaling probabilities and processes to different time-step widths. This was in detail discussed in Section 5.1.1 to be a partially unsolvable problem. Nevertheless, it is sometimes necessary to rephrase a model with a finer or rougher time-resolution to receive more flexibility. Moreover, the analysis of the time-discrete version of the simple population model in Section 5.1.3 showed that a smaller time-step length is also beneficial with respect to stochastic fluctuations.

- Logical ordering of events that occur in one time step. This problem was discussed in Section 5.6.1 in the agent-based GEPOC model, as it was necessary to be aware of the difference between a sub-model that dies before it recreates or vice versa. In this example the processes cannot be scheduled manually as they might occur in any order in reality. In

other examples a manually defined order among the processes might be unavoidable – in the most obvious case: a sub-model cannot leave before it entered.

Time-continuous microscopic models are models that are updated based on event driven time-steps. Hereby the model dynamics are always based on the following iterative process:

1. Evolve time to the next scheduled state-event.
2. This event is now considered to influence the model at precisely this time instant. Therefore, execute all actions correlated with the event and update the model variables.
3. If the occurrence of the event somehow influenced the future of the model, it is necessary to schedule new events. Occasionally delete other planned events that became invalid.
4. Find out which event occurs next and continue with 1.

As microscopic models consist of a large number of sub-models, also a huge number of events have to be handled by the discrete event simulator. It is clear that most events that take place somehow "belong" to specific sub-models. For example, each agent *plans* its own specific death-date, each particle *plans* the time for its next movement, each entity *plans* the time when it arrives at the next waiting queue. This feature poses for a quite natural way to divide and conquer the huge event-list in event-driven microscopic models and can be used to improve performance.

In general the usage of event-based time update is preferred to discrete-time update as both mentioned challenges for time-discrete microscopic models are solved at once. Nevertheless, new challenges occur that make the use of this concept quite tricky.

- Interaction of sub-models with global consequences. Often specific events in the model have an impact not only on one sub-model, but on a larger number of other sub-models. In the easiest case we may think of a contact-event between two sub-models which, clearly, not only influences the future events of one, but of both sub-models. Consequently, heavy problems occur when an occurring event has a global influence on all other sub-models and, hence, on all other scheduled future events. Imagine a fictional agent-based stock market model wherein agents buy and sell stocks – the higher the price, the lower the motivation to buy. Whenever an agent buys, the price of the stock and equivalently the buying-rate of all other agents changes. Consequently, all planned "buying-dates" of all agents have to be cancelled and rescheduled whenever a stock is bought.
- Time dependent parameters. In case a parameter function effects the likelihood of the occurrence of a specific event in a time-dependent way it is not possible to sample the occurrence times of this event by simple exponentially distributed variables. There are essentially two options to solve this problem: First, it is possible to investigate the complete graph of the parameter function and derive a probability distribution for sampling, that suits the given parameter function. Second, the parameter function is discretised to be constant for certain time intervals. Moreover, as soon as one of the discretisation intervals has passed the corresponding event-time is cancelled and re-sampled with a new value.

The second option essentially corresponds to a time-discrete model update, the first option is difficult to derive and has bad performance, as sampling of random variables with arbitrary distribution is a computationally demanding operation.

- Performance of the model. Especially the two aforementioned problems lead to a very computationally intensive simulation that might pose a challenge to any computer as soon as the number of sub-models is large.

- Continuous states. We have discussed in the course of the random-walk model analysis in Section 5.5.1 that it is generally impossible to simulate continuous updates of continuous states with an event-driven mechanism. Hence, either a discretisation needs to be applied or the time-enhancement is modelled with a different time-continuous concept, e.g. a differential equation. The latter is quite usual in simulation of traffic flow (e.g. the famous car-following *Gipps model* [Gipps, 1981]).

Depending on the complexity of the model and the number of sub-models it might even be impossible to define the model in a event-driven style with reasonable effort. In the simple population model 1.1 discrete event concepts could easily be applied successfully to rephrase the model to the time-continuous formulation 1.2 as the model parameters were assumed to be constant. For the agent-based GEPOC model 6.1 this process would be a lot more difficult.

Finally, the choice of terminology for this classification parameter is not unusual. The term can quite often be found in literature reviews to distinguish different modelling approaches. We e.g. find this distinction in reviews about traffic flow models [Treiber et al., 2000], but also describing population/health-care models [Spielauer, 2007].

6.1.2 Classification with Respect to State-Space

We briefly described in Section 4.1 that it is inevitable to formalise the models to be able to perform analysis of microscopic models in general. We introduced the term *formalised microscopic model* and stated that its formulation is the key to apply the mean-field theorems and methods in Sections 4.4, 4.3 and 4.5. The classification parameter presented in this section refers to one specific aspect of this formalisation concept that can be detected in every microscopic model: the **state-space**.

Abbreviated by the symbol Γ throughout this work, we denoted the space that contains the summary of all possible states any sub-model in the microscopic model may have at any time during the simulation. In this specific case we use the term *state* to refer to anything that characterises a specific sub-model at the investigated time-instant and distinguishes it from any other sub-model. That includes for example

- dynamic states (e.g. age 6.1, health-state 5.4 or position 5.1),

- static states, sometimes so-called properties or different sub-model types (e.g. male/female 6.1 or predator/prey in a Lotka-Volterra model [Lotka, 1956]),

- memory of prior states (e.g. memorising the time since last food consumption as required in the WaTor model [Durrett and Levin, 2000]), and

- variables that describe the behaviour of the regarded sub-model (e.g. individual driving behaviour in an (advanced) traffic flow model [Treiber et al., 2000]).

Mathematically spoken, in case any two sub-models I_i and I_j both have state s at a given time-instant t then their future states are (stochastically) equivalent:

$$I_i(t) = s = I_j(t) \Rightarrow \forall A \subset \Gamma, \forall t' \geq t : P(I_i(t') \in A) = P(I_j(t') \in A). \tag{6.1}$$

In the course of this thesis basically three types of state-spaces could be distinguished: discrete, continuous and hybrid, of which the latter was defined as the product-space of the first two. Hereby it is always finite dimensional, as any executable simulation model is parametrised with a finite number of parameters. For this reason, we may also deduce that any state-space of microscopic models in general can bijectively be mapped to one these three classes.

When investigating a microscopic model with respect to its states, it is often an overhead to take into account the full state-space Γ. This is not only because the state-space is too large for analysis, but primarily because not every variable (i.e. dimension) of the state-space is really measured and defined as an output of the model. One example for this is given by the SIR lattice-model 2.1. Herein the position of the agent on the grid clearly belongs to the state-space of each individual, but it is not considered as output variable of the model. Thus, we define the **output-space** Γ' of each sub-model which consists of specifically chosen dimensions $I \subseteq \{1, \ldots, d\}$ of the original d-dimensional state-space Γ

$$\Gamma' := \bigtimes_{i \in I} (\Gamma)_i, \tag{6.2}$$

which describes which elements of the state-space are actually focus of interest. While the size of the state-space is a result from the model definition, the output-space is given by the research question or the model objectives, respectively.

State-discrete microscopic models are models in which the common **state-space** of the sub-models is solely discrete.

Output-discrete microscopic models are models in which the common **output-space** of the sub-models is solely discrete. Clearly, any state-discrete model belongs to this class, though it might be possible, that models with hybrid state-space are output-discrete if the continuous parts of the state-space are not considered as model-output. The best examples for this feature are found in classic microsimulation models or, to be precise, Markov models for health technology assessment (HTA, see [Abler et al., 2013]), wherein discrete health-states are observed, while (potentially) continuous patient parameters like age or blood values may have an influence on the transition probabilities between the health-states, but are not focus of interest.

Anyway, in case of state-/output-discrete microscopic models there is a bijective mapping of Γ/Γ' onto a subset of $\mathbb{N}^d$. Hence, it can be identified with it. Moreover, we could not find any simulation model with finite end-time t_{end} for which a discrete space did not imply the finiteness of it as well.

As $\mathbb{N}^d$ is always countable, there exists another bijective mapping to a subset of $\mathbb{N}$. Thus, it is even possible to regard the state- or output-space as one-dimensional. Consequently, the **aggregated model output** of time- and state-/output- discrete microscopic models can always be stored in a two dimensional array. This is a nice feature for communicating simulation results as this can always be done in simple table-sheets.

State-continuous microscopic models are models wherein the common state-space of the sub-models contains at least one continuous dimension.

Output-continuous microscopic models are models wherein the common output-space of the sub-models contains at least one continuous dimension. Clearly, only state-continuous microscopic models can be output-continuous, but do not have to (as discussed in the last property). Especially for output-continuous microscopic models one additional challenge could be detected in the course of this thesis.

- Interpretation of the simulation results. We briefly discussed in Section 5.5 what happens if simulation output of microscopic models is found on a continuous axis. In order to process and document the aggregated output of the model it is not simply possible to count sub-models with equivalent state, but one has to estimate the density by which the sub-models' continuous states are distributed. Hereby we tend to use a discretisation in form of histogram bins which may lead to unrepresentative results if the size of the bins is not adapted to the total number of sub-models in the model, but decided by some other measure – e.g. the research question. Alternatively the possibility of kernel density estimation poses for a high-quality but resource-demanding alternative.

In general, the following implications hold:

$$\text{state discrete} \;\Rightarrow\; \text{output discrete},$$

$$\text{output continuous} \;\Rightarrow\; \text{state continuous}.$$

Very likely, though, a state-discrete/continuous model will also result in a discrete/continuous output.

Finally, output of microscopic models is not necessarily defined specifically, especially for "more academic" models for research purposes only. In that cases only the state-space of the model can be classified.

6.1.3 Classification with Respect to Randomness

The distinction between stochastic and deterministic models is one of the most strictly defined ones in the whole science. A model is called **deterministic** if the simulation result is uniquely defined by the simulation input and parameters. Otherwise it is called stochastic. Yet, in terms of microscopic models it might be useful to make an additional distinction among the stochastic models as it has a considerable effect on the behaviour of the simulation.

Initial-value-stochastic microscopic models denote models that use randomness to establish the initial states of the model. This feature can often be observed in microscopic models wherein initial values are distributed according to probability densities. In this work most of the microscopic models belong to this class (e.g. the lattice SIR model 5.4 or the agent-based GEPOC model 6.1).

Initial-value-deterministic microscopic models are all models, wherein the initial states of the sub-models are exactly predefined. Any model that does not belong to the prior class belongs to this one and vice versa. One example for such a model is the groundwater pollution random-walk model 5.1.

As we could not find any initial-value-stochastic microscopic models in which the initial distribution of sub-models is not sampled as an iid (independent, identically distributed) sequence of random numbers (probably mixed with deterministic rules), the law of large numbers ensures that the normed aggregated numbers of the initial distribution converges towards the mean-value. Thus, the influence of randomness on the aggregated initial distribution is decreasing with the total number of sub-models.

Update-stochastic microscopic models denote models that use randomness in their update rules. Again, most of the models discussed belong to this type, e.g. the lattice SIR model 5.4 or the agent-based GEPOC model 6.1. Note, that randomness can occur as a transition probability in time-discrete microscopic models or as sampled random transition dates in time-continuous microscopic models.

Update-deterministic microscopic models denote models, wherein the model dynamics are uniquely defined by the initial distribution. In this work only the Game of Life 4.1 belongs to this class.

The following challenge with respect to deterministic or at least update-deterministic microscopic models can be seen.

- As a result of all performed case studies we clearly deduce that a lack of randomness, especially for the update rules, reduces the chances of finding a valid mean-field approximation as deterministic models tend to be more sensitive with respect to local interactions and pattern formation. Hence, it is questionable if the output of completely deterministic (or complex update-deterministic) microscopic models can be used for generating valid **quantitative simulation** results. After intensive literature research, the idea of Petri net modelling [Murata, 1989] for the analysis of reachability of complex logistic processes was the only found example for a deterministic microscopic model that resulted in quantitative exploitable output. Yet, it is questionable, though, if Petri nets are compatible with our idea of a microscopic model.

6.1.4 Classification with Respect to Interaction Level

Probably the most important classification concept of this thesis relates to the level of interaction between the microscopic sub-models as it has the most direct consequence on the mean-field

behaviour of the model.

Non-interacting microscopic models, contact-less- or linear- microscopic models, denote models wherein the sub-models have no contact with each other whatsoever. The alternative name *linear microscopic model* was chosen as any valid mean-field approximation of a contact-less model is always a linear differential equation – ordinary- in case the state-space if discrete, partial- or integral- in case continuous states are contained. In the current thesis only state-continuous microscopic models without contacts were discussed, namely the groundwater pollution random-walk model 5.1 and the agent-based GEPOC model 6.1. Examples for output-/state-discrete microscopic models without interaction are the mentioned HTA focussed Markov models [Abler et al., 2013].

Globally-interacting microscopic models, or non-linear microscopic models, denote models wherein the contacts between the individuals occur on a mean-field level or can be approximated validly on the mean-field level. Hence, a mean-field approximation can be found based on the mean-field theorems which has a non-linear shape due to contacts. Hereby especially *linear contact processes*, i.e. the search for one random contact-partner, plays a key role as it always leads to mean-field models with non-linearities of quadratic order. To see this, compare the microscopic Levins model 2.4 with the corresponding ordinary differential equation (ODE) model 2.3) or the age dependent SIR model 5.4 with the McKendrick equation model 5.3. In these models a linear contact occurred as the current state of at least one aggregated number in the model has a linear influence on the transitions of the individual sub-models.

Locally-interacting microscopic models finally denote models wherein the contacts between the individuals are so locally restricted that they cannot be approximated on the mean-field level with reasonable effort. Hence, the behaviour of the aggregated number cannot be dedicated to linear or non-linear behaviour anymore as it may have any shape. The best examples to see this are Steven Wolfram's one dimensional cellular automata which were briefly classified by him in [Wolfram, 2002]. The particular cellular automata *rule 110* and *rule 30* can be seen to generate almost arbitrarily complex behaviour while *rule 90* seems to generate completely chaotic patterns and *rule 184* results in a stable steady-state very soon after starting the simulation update. It is obvious that the border between globally- and locally interacting microscopic models is not sharply defined. Hence, attribution to one of these two classes should not be seen in terms of black or white, but as a grey-scale map.

The following challenges can be stated for globally- and locally- interacting microscopic models.

- The most obvious challenge for models with interaction is related to performance. Finding a suitable contact partner usually raises the computational effort from $\mathcal{O}(N)$ (N stands for the total number of sub-models) to a higher order, often N^2. The most prominent example for this feature is probably the agent-based Boids model [Reynolds, 1987] which should depict the flocking behaviour of birds or fishes was already mentioned in the course of Section 2.3.1. Each individual agent needs to get in contact with its surrounding agents to

adapt its velocity and direction. Hereby the calculation of finding which agents surround the targeted one is an operation with order $\mathcal{O}(N)$ which has to be done N-times – once for each target agent in the model. As a result it is the most computational demanding operation in the course of the simulation and leads to quick increase of computational efforts when N is chosen large. While contact-less microscopic models like the agent-based GEPOC population model 6.1 are executable with up to eight million agents, the Boids model reaches its computational limits already for a few ten-thousand flock members, even if performance-optimised.

- Closely related, also computational parallelisation of interacting microscopic simulation is challenging while it is straight forward for non-interacting models. Surprisingly, this is easier if the interaction level is as local as possible, as it is common e.g. for cellular automata [Spezzano and Talia, 1999].

- We have mentioned several times that interaction of sub-models may lead to problems for time-continuous microscopic models. For more about that, the reader is referred to the corresponding challenge in the course of the time-update classification in Section 6.1.1.

- Finally, the (quantitative) validation of local-interacting microscopic models is challenging as the local contacts might lead to completely unpredictable and sensitive behaviour.

It is important to mention that there are clear differences between the classification parameter proposed in this section and the classification which became prominent by Steven Wolfram's famous book New Kind of Science [Wolfram, 2002] in 2002 – his fist publication that included the classification was actually published decades before that in 1984 [Wolfram, 1984]. He stated that there are, in principle, four types of behaviour of one-dimensional, deterministic cellular automata behaviour, namely

1. convergent to a steady state,
2. oscillating,
3. complex, interesting behaviour, or
4. chaotic behaviour.

His collaborator at the Santa Fe Institute, Chris Langton, summarised these four behavioural types in the famous sketch seen in Figure 6.1. Though this classification of behaviour was established for one-dimensional, deterministic cellular automata it can to some extend be extended to microscopic models in general.

The mentioned difference between the classification strategies lies within the focus of the classification. Wolfram's classification concept relies on classifying the models with respect to their behaviour which is basically unknown until the model is executed. Our classification can be done ad-priori solely by the model definition and does, in principle, not require simulation of the model.

Nevertheless, the class a microscopic model can be attributed to, according to our classification parameter with respect to interaction level, has a strong influence to which of Wolfram's classes the model belongs. While the behaviour of all non-interacting and the majority of globally-interacting microscopic models can only be found in the first two classes convergent and oscillating, only locally-interacting microscopic models may lead to chaotic behaviour. It is, to us, an open question if behaviour of globally-interacting models can be denoted as complex or if the ability to perform a mean-field approximation, i.e. the possibility to find a macroscopic description of the system, already excludes this option by definition of the term *complex*.

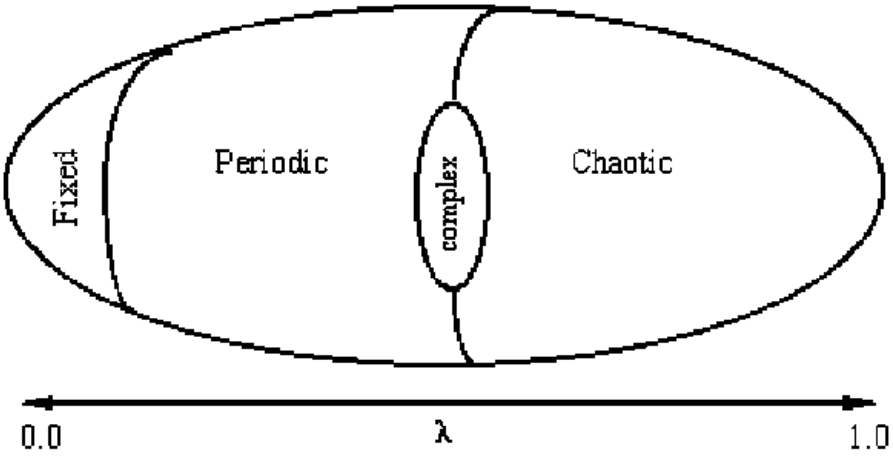


Figure 6.1: Classification of cellular automata according to Steven Wolfram. This sketch was developed by Chris Langton and published in [Langton, 1990]. The one-dimensional measure λ is used to quantify the level fo complexity of the model.

6.2 Classification Summary

In the preceding sections we proposed a classification concept via specific attributive **adjectives** to the word *microscopic models* that can be used to characterise these models in a way that observers, which have no idea about the model in general, are able to get a picture of the challenges that occur in the model and get ideas about how the model behaves during simulation (at least on the aggregated level). This classification concept poses an extension to the current speech describing microscopic models via their modelling method. Figure 6.2 shows most microscopic models briefly discussed in this work and how they are classified with respect to the stated concept. A classification of other famous microscopic models can be found in Table 6.1.

With respect to mean-field analysis of the model the classification makes it possible to, on the one hand, get an idea of the chances of finding a valid mean-field model with reasonable efforts and, on the other hand, how a mean-field model may look like.

- **Time Update.** The classification with respect to randomness in principle has no influence on mean-field analysis related issues. Yet, time-continuous models usually tend to be

simpler with respect to interaction between sub-models. Hence, we could imply that time-continuous models are (on the average) more likely to be depicted on the mean-field level than time-discrete models.

- **State-space.** As mentioned, the classification with respect to output-space considers if there are output dimensions of the model which are continuous or if all dimensions are discrete. It is clear, that as soon as there are continuous variables involved in the model output a mean-field model needs to be a partial- and/or integro- differential equation. For output-discrete models a mean-field is usually an ordinary differential equation, but in cases, in which the state-space includes continuous variables that are not regarded as model output – i.e. in case the model is state-continuous with a hybrid state-space – also partial- and/or integro- differential equation might be required for a valid mean-field description.

- **Randomness.** We discussed in Section 5.4 that adding randomness, especially randomness in the model update processes, to microscopic models has a smoothing effect on the simulated outcomes of the model. Hence, it has a positive influence on sensitivity and, therefore, the chances of finding a valid mean-field model.

- **Interaction.** First, non-interacting and globally-interacting models can (almost) always be described on the mean-field by a macroscopic model. In the first case, linear differential equations result. The second case leads to non-linear differential equations. The more individual contacts between sub-models take place on the local level the more complicated it gets to find a mean-field description of the system. Compare with the pair-approximation introduced in Section 5.3.4, high order systems are required to describe the model if interaction leads to very local effects. Finally, at some stage it is not possible to develop these models with reasonable effort anymore (as stated in the course of the advanced case studies of the Lattice Growth Model 3.1).

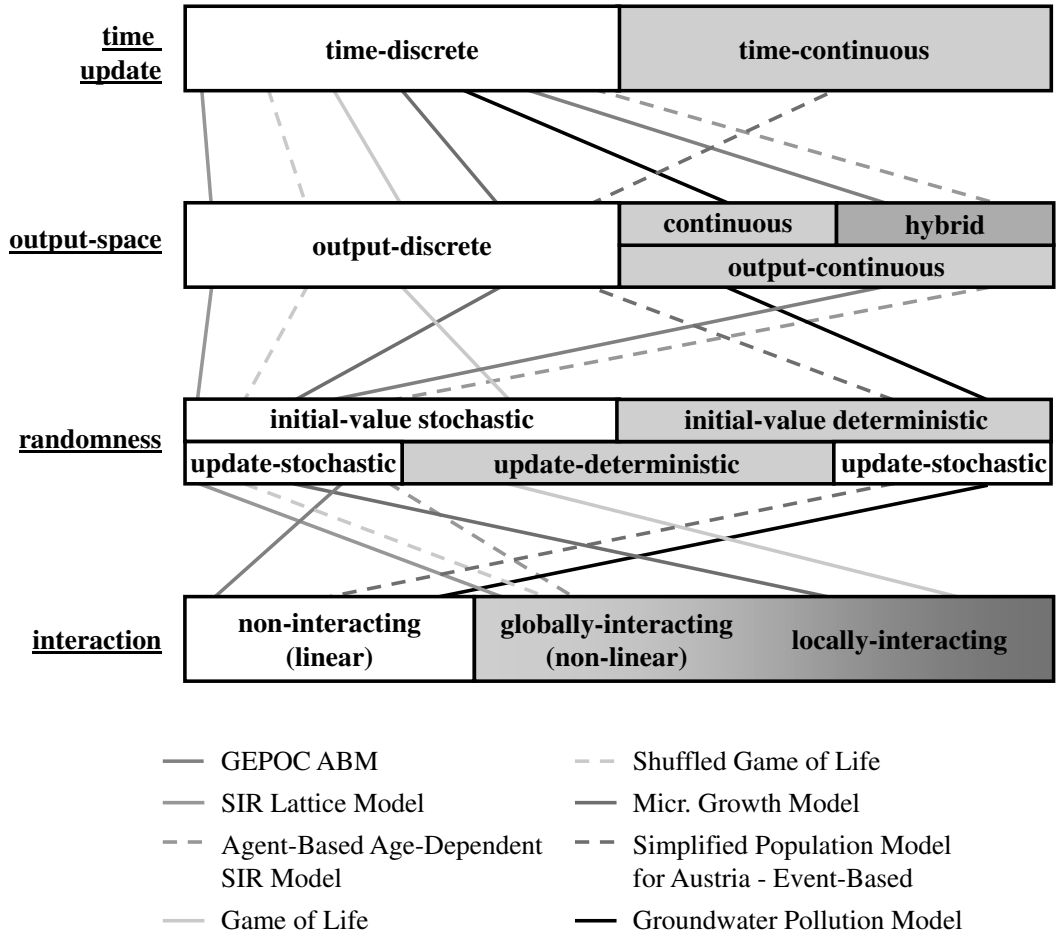


Figure 6.2: Classification of some of the discussed models (GEPOC ABM 6.1, Microscopic Growth Model 3.1, SIR Lattice Model 2.1 and Game of Life4.1) in this thesis with respect to the four defined classification dimensions.

model (application)	time-update	state-space	randomness	interaction
HPP model [Hardy et al., 1976] (fluid flow)	discrete	discrete	initial-value stochastic	globally-interacting
FHP model [Frisch et al., 1986] (fluid flow)	discrete	discrete	completely stochastic	globally-interacting
Ising model [Ising, 1925] (magnetism)	discrete	discrete	initial-value stochastic	globally-interacting
Nagel-Schreckenberg model [Nagel and Schreckenberg, 1992] (traffic flow)	discrete	discrete	update-stochastic	locally-interacting
Gipps model [Gipps, 1981] (traffic flow)	continuous	continuous	deterministic	locally-interacting
Schelling's segregation model [Schelling, 1971] (ghettoism)	discrete	discrete	initial-value stochastic	locally-interacting
Boids model [Reynolds, 1987] (bird flocking)	discrete	continuous	initial-value stochastic	locally-interacting
WaTor [Durrett and Levin, 2000] (predator prey systems)	discrete	discrete	completely stochastic	globally to locally
DYNASIM [Spielauer, 2007] (population based health-care model)	discrete	discrete	completely stochastic	non-interacting
Most other models in [Spielauer, 2007] (same purpose)	discrete	discrete	completely stochastic	non-interacting

Table 6.1: Classification of certain famous models with respect to our defined classes. The models are regarded by their classic definition (i.e. any kind of modification of the model is not considered). As most of the models don't specify the model output we only consider the classification with respect to the model's state-space.

CHAPTER 7

Conclusion and Outlook

In the past chapters we introduced several methods to perform formal analysis of microscopic models focused on the aggregated behaviour of the models. Hereby we established the following contributions.

- First, we summarised and partially extended the state of the art with respect to mean-field analysis focused on models in so-called soft sciences like sociology, epidemiology or economy. This was done by statement of mean-field theorems (MFT) like the general state-space MFT for microscopic models with static (Theorem 3.1) and dynamic (Theorem 4.1) population and the theorem for the macroscopic description of the model's variance (Theorem 5.1). All other mean-field related corollaries in this work follow from these three theorems.
- Moreover, we attempted to make mean-field approximation easier to apply for modellers which are not so familiar with formal stochastic calculus. For this purpose we stated two step-by-step processes (Sections 4.6.2 and 4.6.4), the first to establish a valid macroscopic model given a microscopic one, the second for the inverse purpose. Both guide through mean-field analysis by implicitly applying mentioned mean-field theorems. Main contribution hereby was the derivation of the "correct questions" for these two instructions.
- Based on a number of chosen test-cases we underlined possible fields of direct application for mean-field approximations. Hereby we specifically refer to the ability to test sensitivity of specific microscopic parameters (Section 5.4.3), calibration of specific parameter values (Section 5.6.3) and the acquisition of additional perspectives with respect to, especially continuous, model output (Section 5.5.3 as well as Figures 5.28 and 5.30).
- Finally, a classification of microscopic models was derived to improve documentation capabilities of microscopic models especially with respect to publication, presentation and transfer of model definitions. We are convinced that the four classification parameters improve the quality of describing a model by transferring a clearer picture of the model itself, the challenges involved with it, and its behaviour, at least on the mean-field level.

The current trend in modelling in correlated fields of research goes towards complete disregard of the qualitative as well as quantitative behaviour of the model interpreted as a system by solely using empirical and experimental methods for model development, calibration, verification and validation. Hereby it gets more and more complicated to argue why and how microscopic simulations are superior to (big-) data based approaches like machine-learning, regression modelling or other statistical means (compare with more classical modelling approaches like differential- or difference- equation modelling or stochastic processes). Consequently, decision makers will more and more lose their trust in microscopic simulation models. Thus, tough research questions and therefore also involved microscopic models become growingly complex, we simply cannot afford to blindly accept the fact that these models cannot be analysed with formal (mathematical) means anymore. As the top level goal, the present thesis aimed to motivate modellers to perform formal analysis of their microscopic models as we are convinced that it is necessary. We hope that hereby more and more microscopic models that are now considered as *complex* or *emergent* will, in close future, lose their unexplainable "magic" by finding underlying formal processes that enable their behavioural analysis.

APPENDIX A

Appendix

A.0.1 Proof for the Dynamic General Mean-Field Theorem 4.1

Proof. In order to show the validity of the General Mean-Field Theorem (MFT) for Dynamic Population 4.1 we will define an arbitrary microscopic model fulfilling the prerequisites 4.4.1 and compare it with a correlated model suitable to fulfil the prerequisites 4.3.1 of the classic General MFT for Static Population 3.1. Proving that both models act alike on the aggregate level, we will conclude that the first model fulfils the classic mean-field equation of the second one. Finally, the General MFT for Static Population applied for the second model will be seen to result in the equation stated in the General MFT for Dynamic Population.

We will perform the proof in several steps.

1. First of all an arbitrary population dynamic microscopic model M^1 that fulfils the prerequisites 4.4.1 is taken into account for a limited time interval $[0, t_{\text{end}}]$. Hence, $M^1(t)$ is the state vector of a model consisting of a (large) number of individual agents I_i^1, each assigned a dynamic state in a common state-space Γ^1.

   Based on initially $N_0^1 \in \mathbb{N}$, the number of agents is dynamic and there are rates that agents change their state, are created or destroyed. We summarise according to 4.4.1

   - $\omega : T \times \Gamma \times \Gamma \times \mathbb{R}^n \to \mathbb{R}^+$ – transition-kernel of an agent into a different state.
   - $d : T \times \Gamma \times \mathbb{R}^n \to \mathbb{R}^+$ – death-kernel of an agent.
   - $c : T \times \Gamma \times \Gamma \times \mathbb{R}^n \to \mathbb{R}^+$ – local birth-kernel of an agent triggered by a different agent.
   - $C : T \times \Gamma \times \mathbb{R}^n \to \mathbb{R}^+$ – global birth-kernel of an agent.

2. Before we continue defining a correlated model fulfilling the conditions of the classic mean-field theorem for static population, we introduce a specific measure which is re-

quired for the definition: With

$$\delta_{X(t)}(A) := \sum_{i=1}^{N(t)} \mathbb{1}_A(I_i(t)) = X(A,t) \tag{A.1}$$

a measure that counts the number of agents with state inside the given set A is defined. It really is a measure in the mathematical sense as it is a finite sum of Dirac measures 1.4. As all agents are random processes, so is $\delta_{X(t)}(A)$. Clearly

$$\int_\Gamma \mathbb{1}_A(x) d\delta_{X(t)}(x) = X(A,t)$$

holds.

3. Now, let M^2 denote a model consisting of a static number of individual agents I_i^2. With the notation of the first step of the proof, we define M^2 agents' common state-space by

$$\Gamma^2 := \Gamma^1 \cup \{s_\dagger\},$$

introducing an additional state $s_\dagger$. This state can be imagined as a fictional *idle*-state for not yet created or already destroyed agents.

With a measure μ^1 on Γ^1 we define $\mu^2(A)$ as $\mu^1(A)$ if $s_\dagger \notin A$ and $\mu^1(A \setminus \{s_\dagger\}) + 1$ otherwise (i.e. an extension of the measure μ^1 by the counting measure). Clearly, $T^2 := [0, t_{\text{end}}]$ is defined.

Moreover, we define a transition kernel ω^2 of this population static model based on the transition kernel of the dynamic model. We set

$$\omega^2(t, s, s_1, \vec{x}) := \omega(t, s, s_1, \vec{x}) \quad , s, s_1 \in \Gamma^1, \tag{A.2}$$

$$\omega^2(t, s, s_\dagger, \vec{x}) := d(t, s, \vec{x}), \quad \forall s \in \Gamma^1, \tag{A.3}$$

$$\omega^2(t, s_\dagger, s, \vec{x}) := \frac{\int_{\Gamma^1} c(t, s_1, s, \vec{x}) d\delta_{X^2(t)}(s_1) + C(t, s, \vec{x})}{X(\{s_\dagger\}, t)}, \quad \forall s \in \Gamma^1, \tag{A.4}$$

Note, that $\tilde{X}(s_\dagger, t)$ is finite and equivalent with $X(\{s_\dagger\}, t)$ due to the introduced counting measure, while the density $\tilde{X}(s,t), s \in \Gamma^1$ might not even exist if the probability measure does not fulfil the Radon-Nikodym Theorem 1.1 prerequisites.

Now choose a sufficiently large $N^2 > N_0^1$, say

$$N_2 := LN_0^1 \tag{A.5}$$

for some constant $L > 1$, and initialise that many agents $I_i(0)$ according to the finite density function

$$f^2 : \Gamma^2 \to \mathbb{R}^+ : s \mapsto f^2(s) = \begin{cases} \frac{N_0^1}{N^2} f^1(x), & s \in \Gamma^1 \\ \frac{L-1}{L}, & x = s_\dagger \end{cases}, \tag{A.6}$$

wherein f^1 denotes the initial density of the agents in M^1. Finally, given the initial conditions, model M^2 is fully defined.

4. We analyse if M^2 fulfil the prerequisites 4.3.1 of the population static MFT. We do that using the same bullet points.

   - The time set is an interval by definition.
   - All agents are memoryless processes as all birth- and death- probabilities are defined via rates. The processes are clearly separable and regular, as long as all kernel functions are finite (which is given by the prerequisites of the population dynamic model) and $X(\{s_\dagger\}, t) > 0$. We will deal with the latter problem later on and assume that it is fulfilled now.
   - As all birth- and death- kernels only depend on the state of the agent itself and a finite vector valued function of aggregated numbers, so is ω^2 and hence all transition probabilities. Note, that the dependence on the aggregated numbers has changed now, as the integral (aggregation measure) and the division by $X(s_\dagger, t)$ add additional dependencies on X. Note, that the dependence on $X(s_\dagger, t)$ **only** appears in the birth-kernel.
   - The process is uniquely described by the transition rates

   $$\omega^2(t, s, A) = \int_A \omega^2(t, s, s_1) d\mu(s_1).$$

   - By definition the rates have a kernel representation.
   - By definition the initial states are given by an iid sequence of random numbers with density function f^2.

5. We investigate the resulting mean-field equation according to 3.1. For $s \in \Gamma^1$ we obtain

$$\begin{aligned}
\frac{d}{dt}\varphi(s,t) &= \int_{\Gamma^2} \varphi(s_1,t)\omega(s_1,s,t,\varphi) - \varphi(s,t)\omega(s,s_1,t,\varphi) d\mu^2(s_1) \\
&= \int_{\Gamma^1} \varphi(s_1,t)\omega(s_1,s,t,\varphi) - \varphi(s,t)\omega(s,s_1,t,\varphi) d\mu^2(s_1) \\
&+ \int_{\{s_\dagger\}} \varphi(s_1,t)\omega(s_1,s,t,\varphi) - \varphi(s,t)\omega(s,s_1,t,\varphi) d\mu^2(s_1) \\
&= \int_{\Gamma^1} \varphi(s_1,t)\omega(s_1,s,t,\varphi) - \varphi(s,t)\omega(s,s_1,t,\varphi) d\mu^1(s_1) \\
&\qquad + \underbrace{\varphi(s_\dagger,t)\omega(s_\dagger,s,t,\varphi) - \varphi(s,t)\omega(s,s_\dagger,t,\varphi)}_{=:I}. \qquad \text{(A.7)}
\end{aligned}$$

   We evaluate the corresponding kernel elements of I for $X = \varphi$. As

$$\delta_{\varphi(t)}(A) = \varphi(A,t) = \int_A \varphi(s,t) d\mu^2(s)$$

clearly $\varphi(s,t)$ is a Radon Nikodym density of $\delta_{\varphi(t)}$ w.r. to μ^2. Hence

$$I = \varphi(s_\dagger,t)\left(\frac{\int_{\Gamma^1} c(t,s_1,s,\varphi)d\delta_{\varphi(t)}(s_1)+C(t,s,\varphi)}{\varphi(s_\dagger,t)}\right)-\varphi(s,t)d(t,s,\varphi)$$
$$=\int_{\Gamma^1} c(t,s_1,s,\varphi)\varphi(s_1,t)d\mu^2(s_1)+C(t,s,\varphi)-\varphi(s,t)d(t,s,\varphi).$$

Consequently, the population static mean-field equation for M^2 (limited to Γ^1) is equivalent to the stated equation in the population dynamic model. Also the initial condition

$$\varphi(s,0)=N^2f^2(s)=N^2\frac{N_0^1}{N^2}f^1(s)=N_0^1f^1(s)$$

suits.

6. In order to finally prove that the stated theorem is correct, it remains to show that model M^1 and M^2 (restricted to Γ^1) behave alike on the aggregated level - i.e. that

$$P(X^1(A,t+h)=y(A)|X^1(A,t)=x(A))$$
$$=P(X^2(A,t+h)=y(A)|X^2(A,t)=x(A)) \quad \text{(A.8)}$$

for all natural number -valued measures $x,y:A_\sigma(\Gamma^1)\to\mathbb{N}$, $h>0$ and all $A\in\Gamma^1$.

7. The first (and probably most important) observation is that we only need to prove (A.8) for arbitrarily small h, as both models operate memoryless. Hence, we may choose h small enough to guarantee, that at maximum one state-change has happened in the time interval $[t,t+h]$. Consequently, we only have to consider (A.8) with

   a) $y(A)=x(A)$,

   b) $y(A)=\begin{cases} x(A), & s\notin A \\ x(A)+1, & s\in A \end{cases}$ for some $s\in\Gamma^1$,

   c) $y(A)=\begin{cases} x(A), & s\notin A \\ x(A)-1, & s\in A \end{cases}$ for some $s\in\Gamma^1$, and

   d) $y(A)=\begin{cases} x(A), & s\notin A, s_1\notin A \\ x(A), & s\in A, s_1\in A \\ x(A)-1, & s\in A, s_1\notin A \\ x(A)+1, & s\notin A, s_1\in A \end{cases}$ for some $s,s_1\in\Gamma^1$,

   as only one agent may have changed its state, is born or died during the small time interval $[t,t+h]$. We will only consider b)-d) as the validity of a) follows from the validity of the others by the Theorem of Total Probability.

b) For model M^1 case b) is only possible if one additional agent is born which may be caused by any of the other agents or by the overall system. It calculates to

$$\begin{aligned}
&P(X^1(A,t+h) = y(A) | X^1(A,t) = x(A)) = P(\text{A new agent in } s \text{ is born}) \\
&\quad = \sum_{i=1}^{N^1(t)} P(\text{agent } i \text{ gave birth to an agent with state } s) P(\text{no other events}) \\
&+ P(\text{the system gave birth to an agent with state } s) P(\text{no other events occurred})
\end{aligned}$$

As, by definition, the probability that two or more events occur in a finite time interval is $\mathcal{O}(h^2)$ the probability that "no other events occur"

$$P(\text{no other events occurred}) = 1 - \mathcal{O}(h^2).$$

Moreover,

$$\begin{aligned}
&P(\text{agent } i \text{ gave birth to an agent with state } s) \\
&\qquad = P(\tau^{\text{start}}_{N^1(t)+1} = t+h, I_{N^1(t)+1}(t+h) = s | I_i(t)) \\
&\qquad\qquad = h \cdot c(t, I_i(t), s) + \mathcal{O}(h^2)
\end{aligned}$$

and

$$\begin{aligned}
&P(\text{the system gave birth to an agent with state } s) \\
&\qquad = P(\tau^{\text{start}}_{N^1(t)+1} = t+h, I_{N^1(t)+1}(t+h) = s) \\
&\qquad\qquad = h \cdot C(t,s) + \mathcal{O}(h^2)
\end{aligned}$$

by definition. Thus

$$\begin{aligned}
&P(X^1(A,t+h) = y(A) | X^1(A,t) = x(A)) \\
&\qquad = h \left(\sum_{i=1}^{N^1(t)} c(t, I_i(t), s) + C(t,s) \right) + \mathcal{O}(h^2) \\
&\quad = h \left(\sum_{i=1}^{N^1(t)} \int_{\Gamma^1} c(t, s_1, s) d\delta_{I_i(t)}(s_1) + C(t,s) \right) + \mathcal{O}(h^2) \\
&\qquad = h \left(\int_{\Gamma^1} c(t, s_1, s) d\delta_{X^1(t)}(s_1) + C(t,s) \right) + \mathcal{O}(h^2),
\end{aligned}$$

by definition of the aggregation measure (A.1) based on the Dirac measure 1.4.

The correspondent change of the aggregated number in model M^2 is gained by a state transition from $s_\dagger$ to some state in A. Hence

$$\begin{aligned}
P(X^2(A,t+h) = y(A)|X^2(A,t) = x(A)) \\
&= \sum_{i:I_i(t)=s_\dagger} P(\text{agent } i \text{ switches to } s) \\
&= \sum_{i:I_i(t)=s_\dagger} h \cdot \omega^2(t, s_\dagger, s) + \mathcal{O}(h^2) \\
= X(\{s_\dagger\}, t)h \left(\frac{\int_{\Gamma^1} c(t, s_1, s) d\delta_{X^2(t)}(s_1) + C(t,s)}{X(\{s_\dagger\}, t)} \right) + \mathcal{O}(h^2) \\
&= h \left(\int_{\Gamma^1} c(t, s_1, s) d\delta_{X^2(t)}(s_1) + C(t,s) \right) + \mathcal{O}(h^2).
\end{aligned}$$

Hence

$$\begin{aligned}
P(X^2(A,t+h) = y(A)|X^2(A,t) = x(A)) \\
&= P(X^1(A,t+h) = y(A)|X^1(A,t) = x(A)) + \mathcal{O}(h^2).
\end{aligned}$$

for all $A \in \Gamma^1$. As h was chosen arbitrarily small, the terms corresponding to h^2 can be neglected. Hence, the probabilities for case b) are equivalent in both models.

b) Case c) in model M^1 corresponds to the death of one agent in state s. Consequently,

$$\begin{aligned}
P(X^1(A,t+h) = x(A) - 1|X^1(A,t) = x(A)) = P(\text{An agent with state } s \text{ died}) \\
= \sum_{i:I_i^1(t)\in A} P(\text{agent } i \text{ died}|I_i^1(t) = s) P(\text{no other events occurred}) \\
\sum_{i:I_i^1(t)=s} P(\tau_i^{\text{stop}} = t+h|I_i^1(t) = s) P(\text{no other events occurred}) = \\
= \sum_{i:I_i^1(t)=s} h \cdot d(t,s) + \mathcal{O}(h^2)
\end{aligned}$$

The analogous probability in model M^2 corresponds to a state change from s to $s_\dagger$. It calculates to

$$\begin{aligned}
P(X^2(A,t+h) = y(A)|X^2(A,t) = x(A)) = \\
= \sum_{i:I_i^2(t)=s} P(I_i^2(t+h) = s_\dagger|I_i^2(t) = s) P(\text{no other events occurred}) \\
= \sum_{i:I_i^2(t)=s} h\omega(t,s,s_\dagger) + \mathcal{O}(h^2) \\
= \sum_{i:I_i^2(t)=s} h \cdot d(t,s) + \mathcal{O}(h^2)
\end{aligned}$$

So the two regarded probability measures for case c) are equivalent up to $\mathcal{O}(h^2)$ as well.

b) Case d) follows trivially as the transition rates for a state change within Γ^1 are equivalent in both models.

We have proven that both models behave alike on the aggregated level, in the sense that for any two aggregation measures x, y the transition probability for both models from x to y are equivalent for small h. As both models evolve memoryless, both models also evolve alike on the aggregated level.

Hence, the population-dynamic mean-field equation (4.35) which was valid for model M^2 also holds for model M^1, restricted to Γ^1.

8. Finally the proof of this Theorem is only valid, if the creation rate $c(t, s, A)$ is a finite function which implies that $X(\{s_\dagger\}, t)$ is greater than 0 at all times during the simulation. This leads to two final remarks:

   - The Mean-Field Theorem is valid for finite time sets only. Only this way, $N^1(t) < \infty$ and therefore $N^2 < \infty$ can be assured for all times t which is required for the convergence statement in the population-static theorem.
   - The more agents are created during the simulation process, the higher L with $N^2 = LN_0^1$ has to be chosen. As the error, of the population-static mean-field theorem depends on the total number of (static) agents, also the error of the population dynamic theorem in principle depends on N^2 and, hence, on L. Consequently, the population dynamics model will not converge well towards the mean-field equation if N_0^1 is small while $c(t, s, s_1)$ or $C(t, s)$ is large.

■

A.0.2 Proof for the Discrete-Space Variance Theorem 5.1

Proof. Let

$$\vec{X}(t) := (X(\{s_i\}, t))_{i=1}^{N} := \left(\sum_{j=1}^{N} \mathbb{1}_{\{s_i\}}(I_j(t)) \right)_{i=1}^{N} \tag{A.9}$$

denote the vector-valued aggregated number of a formalised microscopic model with static population according to 1.1 with N agents and state-space

$$\Gamma := \{s_1, \ldots, s_n\} \tag{A.10}$$

that fulfils the prerequisites 4.3.1. Moreover, let

$$\omega(t, i, j, \vec{X}) := \omega(t, s_i, s_j, \vec{u}(\vec{X}, t)) \tag{A.11}$$

denote the agents's transition rates. By statement of the prerequisites there exists a Taylor decomposition of the form

$$\omega(t, i, j, \vec{X}) = \omega(t, i, j, \vec{\phi}) + \frac{(\vec{X} - \vec{\phi})^T}{N} \nabla\omega(t, i, j, \vec{\phi}) + \mathcal{O}\left(\frac{|\vec{X} - \vec{\phi}|^2}{N^2} H(\omega) \right) \tag{A.12}$$

with finite nabla $\nabla\omega$ and Hessian matrix $H(\omega)$, wherein

$$\vec{\phi} := \mathbb{E}(\vec{X}(t)). \tag{A.13}$$

By statement of Theorem 3.1 we already know that $|\vec{X} - \vec{\phi}| = \mathcal{O}(\sqrt{N})$ which makes the first order term of the decomposition

$$\frac{(\vec{X} - \vec{\phi})}{N} \nabla\omega(t, i, j, \vec{\psi}) = \mathcal{O}(N^{-1/2}). \tag{A.14}$$

The second order term is $\mathcal{O}(N^{-1})$. For fixed i, j and $t \in T$ investigate the temporal change of the covariance for small $h > 0$. By linearity of the covariance we receive

$$\begin{aligned}
&\mathrm{Cov}(\vec{X}_i(t+h), \vec{X}_j(t+h)) - \mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t)) \\
&\quad = \mathrm{Cov}(\vec{X}_i(t+h), \vec{X}_j(t+h)) - \mathrm{Cov}(\vec{X}_i(t+h), \vec{X}_j(t)) \\
&\qquad + \mathrm{Cov}(\vec{X}_i(t+h), \vec{X}_j(t)) - \mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t)) \\
&\quad = \mathrm{Cov}(\vec{X}_i(t+h), \vec{X}_j(t+h) - \vec{X}_j(t)) + \mathrm{Cov}(\vec{X}_i(t+h) - \vec{X}_i(t), \vec{X}_j(t)) \\
&\quad = \mathrm{Cov}(\vec{X}_i(t+h), \vec{X}_j(t+h) - \vec{X}_j(t)) - \mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t+h) - \vec{X}_j(t)) \\
&\qquad + \mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t+h) - \vec{X}_j(t)) + \mathrm{Cov}(\vec{X}_i(t+h) - \vec{X}_i(t), \vec{X}_j(t)) \\
&\qquad = \underbrace{\mathrm{Cov}(\vec{X}_i(t+h) - \vec{X}_i(t), \vec{X}_j(t+h) - \vec{X}_j(t))}_{=:I} \\
&\qquad + \underbrace{\mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t+h) - \vec{X}_j(t))}_{=:II} + \underbrace{\mathrm{Cov}(\vec{X}_i(t+h) - \vec{X}_i(t), \vec{X}_j(t))}_{=:III}.
\end{aligned} \tag{A.15}$$

We aim to divide above equation by h and apply the limit $h \to 0$. Hereby an equation for the right-sided differential quotient is derived. As it is easily seen that the same equation follows for the left-sided differential equation, a differential equation system for $\text{Cov}(\vec{X}_i, \vec{X}_j)$ with $i, j \in \{1, \ldots, n\}$ is derived.

1. We initially fix $i \neq j$ and deal with the opposite case later.

$$\begin{aligned} I = \text{Cov}(&\vec{X}_i(t+h) - \vec{X}_i(t), \vec{X}_j(t+h) - \vec{X}_j(t)) \\ &= \underbrace{\mathbb{E}((\vec{X}_i(t+h) - \vec{X}_i(t))(\vec{X}_j(t+h) - \vec{X}_j(t)))}_{=:Ia} \\ &\quad - \underbrace{\mathbb{E}(\vec{X}_i(t+h) - \vec{X}_i(t))\mathbb{E}(\vec{X}_j(t+h) - \vec{X}_j(t))}_{=:Ib} \end{aligned}$$

   By the Theorem of Total Expectation (see 1.5) we may iterate expected values: We assume that the state of the vector is known at t and write

$$Ia = \mathbb{E}\left(\mathbb{E}((\vec{X}_i(t+h) - \vec{X}_i(t))(\vec{X}_j(t+h) - \vec{X}_j(t)))|\vec{X}(t))\right)$$

   As h may be chosen arbitrarily small, we only need to consider the state change of one agent during the observed time-period h. Hence, the aggregated number vector can only be reduced by 1 at one index, while it is increased by 1 at a certain other index. Necessarily the vector at all other indices needs to keep its value. Consequently, the inner expected value can only be non-zero if either one of the agents in state s_i switched to s_j or vice versa (Note, that this is different for $i = j$). In both cases the inner expected value becomes -1

$$Ia = \mathbb{E}\left(-P(\text{switch from } i \text{ to } j|\vec{X}(t)) - P(\text{switch from } j \text{ to } i|\vec{X}(t))\right).$$

   As any of $\vec{X}_j(t)$ agents might switch from i to j the corresponding probability calculates to

$$\begin{aligned} &P(\text{switch from } i \text{ to } j|\vec{X}(t)) \\ &\quad = \vec{X}_i(t) P(I(t+h) = s_j | I(t) = s_i, \vec{X}(t)) P(\text{all others keep their state}). \end{aligned}$$

   Note, that we may set probability that all other agents keep their state to 1 as we consider arbitrarily small h. As

$$\frac{P(I(t+h) = s_j | I(t) = s_i, \vec{X}(t))}{h} \underset{h \to 0}{\longrightarrow} \omega(t, i, j, \vec{X}),$$

   we get

$$\frac{Ia}{h} \underset{h \to 0}{\longrightarrow} -\mathbb{E}(\vec{X}_i(t)\omega(t, i, j, \vec{X})) - \mathbb{E}(\vec{X}_j(t)\omega(t, j, i, \vec{X})). \tag{A.16}$$

   As

$$\omega(t, i, j, \vec{X}) = \omega(t, i, j, \vec{\phi}) + \mathcal{O}(N^{-1/2}),$$

and $\vec{X}_{i/j} = \mathcal{O}(N)$ we get

$$\frac{Ia}{h} \underset{h\to 0}{\longrightarrow} -\mathbb{E}(\vec{X}_i(t))\omega(t,i,j,\vec{\phi}) - \mathbb{E}(\vec{X}_j(t))\omega(t,j,i,\vec{\phi}) + \mathcal{O}(N)\mathcal{O}(N^{-1/2})$$
$$= -\vec{\phi}_i\omega(t,i,j,\vec{\phi}) - \vec{\phi}_j\omega(t,j,i,\vec{\phi}) + \mathcal{O}(\sqrt{N}).$$

Part Ib is very simple to deal with. As both, $\mathbb{E}(\vec{X}_i(t+h) - \vec{X}_i(t))$ and $\mathbb{E}(\vec{X}_j(t+h) - \vec{X}_j(t))$ converge to 0 linearly with h, their product converges quadratic. Hence

$$\frac{Ib}{h} \underset{h\to 0}{\longrightarrow} 0.$$

As a result

$$\frac{I}{h} \underset{h\to 0}{\longrightarrow} -\vec{\phi}_i\omega(t,i,j,\vec{\phi}) - \vec{\phi}_j\omega(t,j,i,\vec{\phi}) + \mathcal{O}(\sqrt{N}). \tag{A.17}$$

2. Clearly, parts II and III can be modified analogously. We start with II. By the Law of Total Covariance (also see 1.5) we receive

$$II = \text{Cov}(\vec{X}_i(t), \vec{X}_j(t+h) - \vec{X}_j(t))$$
$$= \underbrace{\mathbb{E}(\text{Cov}(\vec{X}_i(t), \vec{X}_j(t+h) - \vec{X}_j(t)|\vec{X}(t)))}_{=:IIa}$$
$$+ \underbrace{\text{Cov}(\mathbb{E}(\vec{X}_i(t)|\vec{X}(t))), \mathbb{E}(\vec{X}_j(t+h) - \vec{X}_j(t)|\vec{X}(t)))}_{=:IIb}.$$

Part IIa can easily be seen to be zero as knowing $\vec{X}(t)$ leads to perfect knowledge of $\vec{X}_i(t)$ (compare 1.5). For the same reasons

$$\mathbb{E}(\vec{X}_i(t)|\vec{X}(t)) = \vec{X}_i(t).$$

Hence, $II = IIb$ and can be written to

$$II = \text{Cov}(\vec{X}_i(t), \underbrace{\mathbb{E}(\vec{X}_j(t+h) - \vec{X}_j(t)|\vec{X}(t)))}_{=:IIc}.$$

The value inside the conditional expected value IIc can be $0, 1$ or -1 of which only the latter two cases need to be considered for its calculation. With similar ideas as in Ia we get

$$\mathbb{E}(\vec{X}_j(t+h) - \vec{X}_j(t)|\vec{X}(t))$$
$$= P(\text{switch from any state to } j|\vec{X}(t)) - P(\text{switch from } j \text{ to any state}|\vec{X}(t))$$
$$= \sum_{k=1,k\neq i}^{n} \vec{X}_k(t) P(I(t+h) = s_j | I(t) = s_k, \vec{X}(t))$$
$$- \sum_{k=1,k\neq i}^{n} \vec{X}_j(t) P(I(t+h) = s_k | I(t) = s_j, \vec{X}(t))$$

and

$$\frac{IIc}{h} \underset{h\to 0}{\longrightarrow} \sum_{k=1,k\neq i}^{n} \vec{X}_k(t)\omega(t,k,j,\vec{X}) - \vec{X}_j(t)\omega(t,j,k,\vec{X}) \\ = \sum_{k=1}^{d} \vec{X}_k(t)\omega(t,k,j,\vec{X}) - \vec{X}_j(t)\omega(t,j,k,\vec{X}).$$

Using the linearity of the covariance this leads to

$$\frac{II}{h} \underset{h\to 0}{\longrightarrow} \sum_{k=1}^{d} \mathrm{Cov}(\vec{X}_i(t), \vec{X}_k(t)\omega(t,k,j,\vec{X})) - \mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t)\omega(t,j,k,\vec{X})). \quad \text{(A.18)}$$

In order to finally close the equation system we need to "drag" the ω parts out of the covariances. Therefore, we use Taylor decomposition of the terms $\vec{X}_k(t)\omega(t,k,j,\vec{X})$ at $\vec{X} = \vec{\phi}$. Note, that by Theorem 3.1 still

- $\vec{\phi} = \mathcal{O}(N)$, $\vec{X} = \mathcal{O}(N)$,
- $\mathbb{V}(X) = \mathcal{O}(N)$, and
- $|\vec{X} - \vec{\phi}| = \mathcal{O}(N^{1/2})$

hold asymptotically. By the product law and the stated Taylor-decomposition of ω we get

$$\begin{aligned}
&\vec{X}_k(t)\omega(t,k,j,\vec{X}) = \left(\vec{\phi}_k(t) + (\vec{X}_k(t) - \vec{\phi}_k(t))\right)\omega(t,k,j,\vec{X}) \\
&= \left(\vec{\phi}_k(t) + (\vec{X}_k(t) - \vec{\phi}_k(t))\right)\left(\omega(t,k,j,\vec{\phi}) + \frac{(\vec{X}(t) - \vec{\phi}(t))}{N}\nabla\omega(t,k,j,\vec{\phi}) + \mathcal{O}(N^{-1})\right) \\
&\qquad = \vec{\phi}_k(t)\omega(t,k,j,\vec{\phi}) \\
&+ (\vec{X}(t) - \vec{\phi}(t))^T \cdot \left((0,\dots,0,\underbrace{1}_{k},0,\dots,0)^T\omega(t,k,j,\vec{\phi}) + \vec{\phi}_k(t)\frac{\nabla(\omega(t,k,j,\vec{\phi}))}{N}\right) \\
&+ (\vec{X}(t) - \vec{\phi}(t))^T \cdot \underbrace{\left((\vec{0},\dots,\underbrace{2\frac{\nabla(\omega(t,k,j,\vec{\phi}))}{N}}_{k},\dots,\vec{0})^T + \mathcal{O}(N^{-1})\right)}_{\mathcal{O}(N^{-1})} \cdot (\vec{X}(t) - \vec{\phi}(t)).
\end{aligned} \quad \text{(A.19)}$$

Clearly, the covariance applied on this term can be used on all summands separately for its multilinearity property.

Applying $\mathrm{Cov}(\vec{X}_k, \cdot)$ on the **second order** parts of the Taylor decomposition we obtain terms in form of

$$\mathrm{Cov}(\vec{X}_k, (\vec{X}_l - \vec{\phi}_l)(\vec{X}_m - \vec{\phi}_m))\mathcal{O}(N^{-1}). \quad \text{(A.20)}$$

Using the Cauchy-Schwartz inequality and the Theorem of Steiner we obtain

$$\begin{aligned}\text{Cov}(\vec{X}_k,(\vec{X}_l-\vec{\phi}_l)(\vec{X}_m-\vec{\phi}_m)) &\leq \sqrt{\mathbb{V}(\vec{X}_k)\mathbb{V}((\vec{X}_l-\vec{\phi}_l)(\vec{X}_m-\vec{\phi}_m))}\\ &= \sqrt{\mathcal{O}(N)\left(\mathbb{E}((\vec{X}_l-\vec{\phi}_l)^2(\vec{X}_m-\vec{\phi}_m)^2)-\mathbb{E}((\vec{X}_l-\vec{\phi}_l)(\vec{X}_m-\vec{\phi}_m))^2\right)}\\ &\leq \sqrt{\mathcal{O}(N)\mathbb{E}((\vec{X}_l-\vec{\phi}_l)^2(\vec{X}_m-\vec{\phi}_m)^2)}\\ &= \sqrt{\mathcal{O}(N)\mathbb{E}\left(\mathcal{O}(\sqrt{N})^2\mathcal{O}(\sqrt{N})^2\right)}\\ &= \sqrt{\mathcal{O}(N)\mathcal{O}(N^2)} = \mathcal{O}(N^{3/2}).\end{aligned}$$

Hence, these terms have asymptotic order $\mathcal{O}(N^{3/2})/N = \mathcal{O}(\sqrt{N})$.

The **zero order** term of the Taylor decomposition vanishes as it is deterministic

$$\text{Cov}(\vec{X}_k,\vec{\phi}_k(t)\omega(t,k,j,\vec{\phi})) = \text{Cov}(\vec{X}_k,1)\vec{\phi}_k(t)\omega(t,k,j,\vec{\phi}) = 0. \tag{A.21}$$

Hence, only the **first order** terms of the Taylor decomposition remain. As $\text{Cov}(\vec{X}_k,\vec{X}-\vec{\phi}) = \text{Cov}(\vec{X}_k,\vec{X}) - \text{Cov}(\vec{X}_k,\vec{\phi}) = \text{Cov}(\vec{X}_k,\vec{X}) - 0$ the first order terms can be calculated to

$$\begin{aligned}&\text{Cov}(\vec{X}_i(t),\vec{X}_k(t)\omega(t,k,j,\vec{X}))\\ &\qquad= \text{Cov}(\vec{X}_i(t),\vec{X}_k(t))\omega(t,k,j,\vec{\phi})\\ &\qquad+\sum_{m=1}^{n}\text{Cov}(\vec{X}_i(t),\vec{X}_m(t))\frac{\vec{\phi}_k(t)}{N}\frac{\partial\omega(t,k,j,\vec{\phi})}{\partial\frac{\vec{X}_m}{N}}+\mathcal{O}(\sqrt{N}).\end{aligned}$$

Consequently

$$\begin{aligned}&\frac{II}{h}\underset{h\to 0}{\longrightarrow}\sum_{k=1}^{d}\text{Cov}(\vec{X}_i,\vec{X}_k\omega(t,k,j,\vec{X}))-\text{Cov}(\vec{X}_i,\vec{X}_j\omega(t,j,k,\vec{X}))\\ &\qquad=\sum_{k=1}^{d}\text{Cov}(\vec{X}_i,\vec{X}_k)\omega(t,k,j,\vec{\phi})-\text{Cov}(\vec{X}_i,\vec{X}_j)\omega(t,j,k,\vec{\phi})\\ &+\sum_{m=1}^{d}\text{Cov}(\vec{X}_i(t),\vec{X}_m(t))\sum_{k=1}^{d}\left(\frac{\vec{\phi}_k(t)}{N}\frac{\partial\omega(t,k,j,\vec{\phi})}{\partial\frac{\vec{X}_m}{N}}-\frac{\vec{\phi}_j(t)}{N}\frac{\partial\omega(t,j,k,\vec{\phi})}{\partial\frac{\vec{X}_m}{N}}\right)+\mathcal{O}(\sqrt{N}).\end{aligned} \tag{A.22}$$

Analogously for part III we receive

$$\begin{aligned}\frac{III}{h} &\xrightarrow[h\to 0]{} \sum_{k=1}^{d} \mathrm{Cov}(\vec{X}_j, \vec{X}_k\omega(t,k,i,\vec{X})) - \mathrm{Cov}(\vec{X}_j, \vec{X}_i\omega(t,i,k,\vec{X})) \\ &= \sum_{k=1}^{d} \mathrm{Cov}(\vec{X}_j, \vec{X}_k)\omega(t,k,i,\vec{\phi}) - \mathrm{Cov}(\vec{X}_j, \vec{X}_i)\omega(t,i,k,\vec{\phi}) \\ +\sum_{m=1}^{d} \mathrm{Cov}(\vec{X}_j(t), \vec{X}_m(t)) &\sum_{k=1}^{d} \left(\frac{\vec{\phi}_k(t)}{N} \frac{\partial\omega(t,k,i,\vec{\phi})}{\partial \frac{\vec{X}_m}{N}} - \frac{\vec{\phi}_i(t)}{N} \frac{\partial\omega(t,i,k,\vec{\phi})}{\partial \frac{\vec{X}_m}{N}} \right) + \mathcal{O}(\sqrt{N})\end{aligned} \tag{A.23}$$

3. Finally putting I, II and III together we obtain the equation

$$\begin{aligned}\lim_{h\to\infty} &\frac{\mathrm{Cov}(\vec{X}_i(t+h), \vec{X}_j(t+h)) - \mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t))}{h} \\ &= -\vec{\phi}_i(t)\omega(t,i,j,\vec{\phi}) - \vec{\phi}_j(t)\omega(t,j,i,\vec{\phi}) \\ &+ \sum_{k=1}^{d} \mathrm{Cov}(\vec{X}_i(t), \vec{X}_k(t))\omega(t,k,j,\vec{\phi}) - \mathrm{Cov}(\vec{X}_i(t), \vec{X}_j(t))\omega(t,j,k,\vec{\phi}) \\ &+ \sum_{k=1}^{d} \mathrm{Cov}(\vec{X}_j(t), \vec{X}_k(t))\omega(t,k,i,\vec{\phi}) - \mathrm{Cov}(\vec{X}_j(t), \vec{X}_i(t))\omega(t,i,k,\vec{\phi}) \\ &+ \sum_{m=1}^{d} \mathrm{Cov}(\vec{X}_i(t), \vec{X}_m(t)) \sum_{k=1}^{d} \left(\frac{\vec{\phi}_k(t)}{N} \frac{\partial\omega(t,k,j,\vec{\phi})}{\partial \frac{\vec{X}_m}{N}} - \frac{\vec{\phi}_j(t)}{N} \frac{\partial\omega(t,j,k,\vec{\phi})}{\partial \frac{\vec{X}_m}{N}} \right) \\ &+ \sum_{m=1}^{d} \mathrm{Cov}(\vec{X}_j(t), \vec{X}_m(t)) \sum_{k=1}^{d} \left(\frac{\vec{\phi}_k(t)}{N} \frac{\partial\omega(t,k,i,\vec{\phi})}{\partial \frac{\vec{X}_m}{N}} - \frac{\vec{\phi}_i(t)}{N} \frac{\partial\omega(t,i,k,\vec{\phi})}{\partial \frac{\vec{X}_m}{N}} \right) \\ &+ \mathcal{O}(\sqrt{N}).\end{aligned} \tag{A.24}$$

Clearly, $\lim_{h\to\infty} \frac{\mathrm{Cov}(\vec{X}_i(t),\vec{X}_j(t)) - \mathrm{Cov}(\vec{X}_i(t-h),\vec{X}_j(t-h))}{h}$ leads to the same equations which makes $\vec{X}$ steadily differentiable. Hence, we found a differential equation for $\mathrm{Cov}(\vec{X}_i, \vec{X}_j)$ which is fulfilled up to an error of $\sqrt{N}$. As the solution $\vec{\varphi}(t)$ of the mean-field equation 4.42 approximates $\vec{\phi}(t) = \mathbb{E}(\vec{X}(t))$ with the same asymptotic error, we can replace $\vec{\phi}$ in (A.24) by $\vec{\varphi}$ without increasing the error order. As solutions of differential equations depend steadily on the equation parameters (compare with the Lemma of Gronwall) any

solution $\xi_{i,j}(t)$ of the system

$$\begin{aligned}\frac{d}{dt}\xi_{i,j}(t) = &-\vec{\varphi}_i(t)\omega(t,i,j,\vec{\varphi}) - \vec{\varphi}_j(t)\omega(t,j,i,\vec{\varphi}) \\ &+ \sum_{k=1}^{d} \xi_{i,k}(t)\omega(t,k,j,\vec{\varphi}(t)) - \xi_{i,j}(t)\omega(t,j,k,\vec{\varphi}(t)) \\ &+ \sum_{k=1}^{d} \xi_{j,k}(t)\omega(t,k,i,\vec{\varphi}(t)) - \xi_{j,i}(t)\omega(t,i,k,\vec{\varphi}(t)) \\ &+ \sum_{m=1}^{d} \xi_{i,m}(t) \sum_{k=1}^{d} \left(\frac{\vec{\varphi}_k(t)}{N} \frac{\partial\omega(t,k,j,\vec{\varphi})}{\partial \frac{\vec{X}_m}{N}} - \frac{\vec{\varphi}_j(t)}{N} \frac{\partial\omega(t,j,k,\vec{\varphi})}{\partial \frac{\vec{X}_m}{N}} \right) \\ &+ \sum_{m=1}^{d} \xi_{j,m}(t) \sum_{k=1}^{d} \left(\frac{\vec{\varphi}_k(t)}{N} \frac{\partial\omega(t,k,i,\vec{\varphi})}{\partial \frac{\vec{X}_m}{N}} - \frac{\vec{\varphi}_i(t)}{N} \frac{\partial\omega(t,i,k,\vec{\varphi})}{\partial \frac{\vec{X}_m}{N}} \right) \end{aligned} \tag{A.25}$$

with $\xi_{i,j}(0) = \text{Cov}(\vec{X}_i(0), \vec{X}_j(0))$ approximates the covariance curve $\text{Cov}(\vec{X}_i(t), \vec{X}_j(t))$ with error depending on $\sqrt{N}$. As $\mathbb{V}(\vec{X}) = \mathcal{O}(N)$ this error comparably vanishes on the scale of the (co-)variance.

4. It remains to discuss the case $i = j$. Hereby only part I is going to change as the expected value in

$$\begin{aligned} I = \mathbb{E}\left(\mathbb{E}((\vec{X}_i(t+h) - \vec{X}_i(t))(\vec{X}_j(t+h) - \vec{X}_j(t)))|\vec{X}(t))\right) \\ = \mathbb{E}\left(\mathbb{E}((\vec{X}_i(t+h) - \vec{X}_i(t))(\vec{X}_i(t+h) - \vec{X}_i(t)))|\vec{X}(t))\right) \end{aligned}$$

can either be 1 if an agent switched from any other state to state s_i or $(-1)^2 = 1$ if an agent switched from s_i to any other state. Consequently, we obtain

$$\begin{aligned} &\mathbb{E}((\vec{X}_i(t+h) - \vec{X}_i(t))(\vec{X}_i(t+h) - \vec{X}_i(t)))|\vec{X}(t)) \\ &\quad = \sum_{k=1,k\neq j}^{n} \vec{X}_k(t) P(I(t+h) = s_i | I(t) = s_k, \vec{X}(t)) \\ &\qquad + \sum_{k=1,k\neq j}^{n} \vec{X}_i(t) P(I(t+h) = s_k | I(t) = s_i, \vec{X}(t)) \end{aligned}$$

and

$$\begin{aligned} &\frac{\mathbb{E}((\vec{X}_i(t+h) - \vec{X}_i(t))(\vec{X}_i(t+h) - \vec{X}_i(t)))|\vec{X}(t))}{h} \xrightarrow[h\to 0]{} \\ &\quad = \sum_{k=1,k\neq i}^{n} \vec{X}_k(t)\omega(t,k,i,\vec{X}) + \vec{X}_i(t)\omega(t,i,k,\vec{X}) \\ &\quad = \sum_{k=1}^{n} \left(\vec{X}_k(t)\omega(t,k,i,\vec{X}) + \vec{X}_i(t)\omega(t,i,k,\vec{X})\right) - 2\vec{X}_i(t)\omega(t,i,i,\vec{X}). \end{aligned}$$

Applying the Taylor approximation we obtain

$$\frac{I}{h} \underset{h\to 0}{\longrightarrow} \sum_{k=1}^{n} \left(\vec{\phi}_k(t)\omega(t,k,i,\vec{\phi}) + \vec{\phi}_i(t)\omega(t,i,k,\vec{X})\right) - 2\vec{\phi}_i(t)\omega(t,i,i,\vec{\phi}).$$

The right part of this term matches part I of the $i \neq j$ case applied to $i = j$. For this reason we may finally write the complete differential equation system using an indicator function:

$$\begin{aligned}
\frac{d}{dt}\xi_{i,j}(t) = &-\vec{\varphi}_i(t)\omega(t,i,j,\vec{\varphi}) - \vec{\varphi}_j(t)\omega(t,j,i,\vec{\varphi}) \\
&+ \mathbb{1}_i(j)\sum_{k=1}^{n} \vec{\varphi}_k(t)\omega(t,k,i,\vec{\varphi}(t)) + \vec{\varphi}_i(t)\omega(t,i,k,\vec{\varphi}(t)) \\
&+ \sum_{k=1}^{d} \xi_{i,k}(t)\omega(t,k,j,\vec{\varphi}(t)) - \xi_{i,j}(t)\omega(t,j,k,\vec{\varphi}(t)) \\
&+ \sum_{k=1}^{d} \xi_{j,k}(t)\omega(t,k,i,\vec{\varphi}(t)) - \xi_{j,i}(t)\omega(t,i,k,\vec{\varphi}(t)) \\
&+ \sum_{m=1}^{d} \xi_{i,m}(t) \sum_{k=1}^{d} \left(\frac{\vec{\varphi}_k(t)}{N} \frac{\partial\omega(t,k,j,\vec{\varphi})}{\partial \frac{\vec{X}_m}{N}} - \frac{\vec{\varphi}_j(t)}{N} \frac{\partial\omega(t,j,k,\vec{\varphi})}{\partial \frac{\vec{X}_m}{N}} \right) \\
&+ \sum_{m=1}^{d} \xi_{j,m}(t) \sum_{k=1}^{d} \left(\frac{\vec{\varphi}_k(t)}{N} \frac{\partial\omega(t,k,i,\vec{\varphi})}{\partial \frac{\vec{X}_m}{N}} - \frac{\vec{\varphi}_i(t)}{N} \frac{\partial\omega(t,i,k,\vec{\varphi})}{\partial \frac{\vec{X}_m}{N}} \right) \qquad \text{(A.26)}
\end{aligned}$$

5. Finally we deal with the initial condition. Let $\vec{f}$ denote the distribution of the initial states of the agents which are independent by definition, then

$$\begin{aligned}
\mathrm{Cov}(\vec{X}_i(0), \vec{X}_j(0)) &= \mathrm{Cov}\left(\sum_{k=1}^{N} \mathbb{1}_i(I_k(0)), \sum_{l=1}^{N} \mathbb{1}_j(I_l(0))\right) \\
&= \sum_{k=1}^{N}\sum_{l=1}^{N} \mathrm{Cov}\left(\mathbb{1}_i(I_k(0)), \mathbb{1}_j(I_l(0))\right) = \sum_{k=1}^{N} \mathrm{Cov}\left(\mathbb{1}_i(I_k(0)), \mathbb{1}_j(I_k(0))\right) \\
&= \sum_{k=1}^{N} \mathbb{E}\left(\mathbb{1}_i(I_k(0))\mathbb{1}_j(I_k(0))\right) - \mathbb{E}(\mathbb{1}_i(I_k(0)))\mathbb{E}(\mathbb{1}_j(I_k(0))) \\
&= \sum_{k=1}^{N} \mathbb{E}\left(\mathbb{1}_i(I_k(0))\mathbb{1}_j(I_k(0))\right) - P(I_k(0) = s_i)P(I_k(0) = s_j) \\
&= \sum_{k=1}^{N} \mathbb{E}\left(\mathbb{1}_i(I_k(0))\mathbb{1}_j(I_k(0))\right) - \vec{f}_i\vec{f}_j.
\end{aligned}$$

For $i \neq j$ the product in the first expected value can never be non-zero as $I_k(0)$ cannot have both states s_i and s_j. For $i = j$ the expected value equals the probability that $I_k(0) = s_i$ which is equivalent with $\vec{f}_i$ by definition. As, finally, the summands do not depend on the index anymore, it resolves to a multiplication with N. Thus

$$\text{Cov}(\vec{X}_i(0), \vec{X}_j(0)) = \mathbb{1}_i(j) N f_i - f_i f_j. \tag{A.27}$$

■

List of Figures

2.1 Illustration of Monte Carlo Integration on the example of $\sin(x)e^{-x^2}$. Though the function cannot be integrated analytically, the integral can be approximated by the quotient of the number of points below the curve and the number of points in total. 9

2.2 Illustration of Buffon's Needle Problem with $a = 1$ and $b = 0.5$ with $N = 1000$ needles, randomly dropped onto the grid. They are marked blue if they intersect with a lines, red otherwise. 10

2.3 Illustration of a cellular automaton (Ising Model) before (left) and after the update (right). Cells are arranged on a rectangular grid and their states are marked with one of two possible colours. The neighbourhood used considers all 4 directly connected cells and the cell itself. 14

2.4 Sketch of Schelling's Segregation model depicted as cellular automaton (left) and as agent-based model. The subjects of the model are coloured with grey background. 16

2.5 Minimalist example of a bouncing ball described by an Event Graph. The "bounce" event occurs every time, the ball hits the floor. In case of a bounce the velocity of the ball at the time instant of the bounce is reduced by a damping factor $d < 1$. The time to the next bounce can easily be calculated by the larger root of the equation $0 = \frac{-gt^2}{2} + vt$ 23

2.6 Event Graph of a single-server queue. Variable Q, initially 0, describes the length of the queue, $S \in \{True, False\}$ describes whether or not the server is available. The time between two arrivals of customers t_a and the time required to serve one of them t_s can be fixed with values or positive random numbers. 23

2.7 Single-server queue modelled with the Enterprise Library of AnyLogic 7. The underlying model is equivalent to the one shown in the event graph in Figure 2.6, yet the process is more intuitively depicted here. 24

2.8 Microscopic model (agent-based model) implemented in an object-oriented programming language (left) compared to the same implementation in a procedural language (right) 28

2.9 Screenshots of a 3-d Boids simulation run with 4000 individual swarm members. While the single sub-models only follow three simple rules, the swarm, as a whole, moves almost unpredictable. 33

2.10 Sample solution of the Lorenz attractor for $\sigma = 10, \beta = 3, \rho = 28$ in three dimensions x, y, z. The curve runs from blue $t = 0$ to red $t = 100$. 34

3.1 Comparison of different methods for density estimation with 100 uniformly $U(0,1)$ generated pseudo random numbers. While the upper two images show standard kernel density estimations with Gaussian and uniform kernel the third plot shows a standard histogram with a comparable number of bins. 55

4.1 Role of the formalised microscopic model for mean-field analysis. 74
4.2 General concept of the application of mean-field theorems. This image is originated in [Bicher et al., 2017a] 80

5.1 Comparison of the two possible formulas for scaling a model from time-step length 1 to time-step length 0.5. 128
5.2 Histogram plots of different time-scaling scenarios. The model with step-size 1 is shown in the top three plots, the corresponding model with step-size 0.5 is shown below. Formula (5.10) was used to calculate $p_{0.5}$ from p_1. 128
5.3 Histogram plots comparing discrete-time model M_1 and continuous-time model M_0. Transition probability of $p_1 = 0.2$ was used for the discrete-time model, which was used as rate-parameter in the exponential distribution for the continuous-time model. 130
5.4 The pictures show a comparison between the described event-driven (blue) and discrete time-step (red) population model with different step-sizes dt from 0.02 to 2. The broad, solid lines mark the sample mean of 100 simulation-runs. The dashed lines mark curves which are, each, one sample standard deviation away from the mean. The thin solid lines, finally, show some sample paths of the model. 134
5.5 Alternative representation possibility of the event-driven population model via the Event Graph. The three event-nodes at the right hand side are initialised with the index parameter i. Hence, they only cancel and reschedule themselves according to this parameter. 135
5.6 Diagram of the time evolution in the microscopic SIR model. This diagram has already been published in [Bicher and Popper, 2013]. 137
5.7 Comparison of sample runs of the described SIR ABM and the two derived mean-field models. The left part of the figure shows 10 sample runs of the ABM and their sample mean. The right part shows a comparison of the sample mean with the solution of the classic and the cubic SIR model. 141
5.8 Comparison of sample runs of the described SIR ABM and the two derived mean-field models with higher density and infection parameter λ than in Figure 5.7. . . . 142
5.9 Comparison of sample runs of the described SIR ABM and the two derived mean-field models with the same parameters as for Figure 5.7. Yet the movement phase was skipped. 143
5.10 Comparison of sample variance of 500 runs of the Lattice SIR model with 100×100 sites (left) and the numerically generated curves of the differential equation system (5.24)-(5.26) (right). 144
5.11 Sketch of the established microscopic Levins Model. While the population is within its resource-bounds (upper sketch), birth processes are possible. If the population is too large, the overshoot agents have an increased death probability. 148

5.12 Levins ODE Model with immigration (5.27) in comparison with the defined agent-based model, Model 2.4, using different parameter settings. The upper-left picture shows that the simplification done in (5.34) disturbs the similarity of the two approaches if the population is too much overcrowded. Here, the microscopic model result drift off slightly. 149
5.13 Circles with specific radius for all three specified lattice-metrics, with $c_{i,j}$ being the centred site. For radius 1, the $d_1(\cdot)$ metric and the $d_2(\cdot)$ metric lead to the Von Neumann neighbourhood, the last one $d_\infty(\cdot)$ results in the Moore neighbourhood. . 152
5.14 Sample result of the microscopic growth model 3.1 for different neighbourhood concepts. A grid with $N_1 = N_2 = 100$ sites was used. Hence, the fluctuations of these curves are expected to have asymptotic size of $\mathcal{O}(\sqrt{10000}) = \mathcal{O}(100)$ and can be neglected. 155
5.15 Sample plot as Figure 5.14 included the 160
5.16 Comparison of the results of the one-sided 1-d neighbourhood growth model (Model 3.1, Scenario e) and the results of the derived ODE recursion (5.45). The counted quantities of $[0k]$ combinations for k from 1 to 20 in the microscopic model are shown to the left, the corresponding ODE solution curve is shown to the right. . . . 161
5.17 Solid lines show model scenarios with different neighbourhoods. Although all of them use a neighbourhood with four sites, the results differ a lot. The pair approximation is sketched with circles while the dotted lines show the limit cases for the logistic growth and the linear growth ODE models. 167
5.18 Some classic Conway's Game of Life patterns 171
5.19 The two plots show the same histogram of the 2000 described experiments, wherein the lower plot is equivalent to the upper one with restricted y-axis. In each of those experiments one additional living site was added to a reference initial condition of the GoL. Moreover, the final states of the modified initial conditions after 100 GoL steps were compared to the final state of the reference initial condition. The outcome of each experiment is the number of sites that have a different state. 172
5.20 Based on a reference initial conditions seen in the upper-middle, two additional living sites were added leading to the initial conditions to the upper-right and the upper-left. 100 update-steps of the GoL CA were simulated and the results are plotted directly below the corresponding initial condition. Only the living site added in the right initial condition had an impact on the outcome. 173
5.21 Fifty GoL results with a 2500 sites automaton and randomly chosen 1000 initially alive sites (black), compared with their mean-value (red) and the numeric solution of the Mean-Field ODE (blue). The right plots show a zoom to the first 10 time-steps. 175
5.22 Fifty GoL results with a 2500 sites automaton and randomly chosen 250 initially alive sites (black), compared with their mean-value (red) and the numeric solution of the Mean-Field ODE (blue). The left plots show a zoom to the first 10 time-steps. 176
5.23 Comparison between the results of the classic GoL, the shuffled GoL and the calculated ODE mean-field model for 30 percent initially living sites, i.e. $X_0 = 750 = 0.3 \cdot 2500$. 179

5.24 Interestingly high standard deviation for initial condition $X_0 = 500$ sites as seen in the left plot results from strange instability of the simulation results as seen in the right plot. About two third of the simulation results lead to an approximately stable population while it quickly dies out in all other runs. 180
5.25 Steady state (ss) analysis of the mean-field ODE (5.75). 180
5.26 Fictional mortality-rate $\mu(a) \equiv \overline{\mu}(a)$ and recovery rate $\gamma(a)$. 188
5.27 Heat-map of the fictional contact-rate $\kappa(a_1, a_2)$. The intended band-matrix structure appears because of parent-children-relationships. 191
5.28 Results for the infected density $I(t)/N$ compared for both models with $N = 8000, dt = 0.1$ and a given initial condition. The parameters functions were used as defined in Figures 5.27 and 5.26. The ABM results can be seen to the right and are plotted as histogram with 40 bins. IPDE results gained with (modified) method of lines are shown to the left. 191
5.29 Visualisation of model results for the susceptible population of the age-dependent SIR model. Results of the IPDE model (Model 5.3) are seen at the top-left. Different density estimations of the microscopic model (Model 5.4) results with $N = 1000$ agents can be seen on the other three plots. 195
5.30 Visualisation of some numerically gained results of the GEPOC IPDE model 6.2. The density of the male population $M(a, t)$ is shown in the upper-left corner, different cross-sections are shown in the smaller plots around it. 204

6.1 Classification of cellular automata according to Steven Wolfram. This sketch was developed by Chris Langton and published in [Langton, 1990]. The one-dimensional measure λ is used to quantify the level fo complexity of the model. 220
6.2 Classification of some of the discussed models (GEPOC ABM 6.1, Microscopic Growth Model 3.1, SIR Lattice Model 2.1 and Game of Life4.1) in this thesis with respect to the four defined classification dimensions. 222

Bibliography

[noa, a] AnyLogic, https://www.anylogic.com/.

[noa, b] Arena, https://www.arenasimulation.com/.

[noa, c] Enterprise Dynamics, https://www.incontrolsim.com/product/enterprise-dynamics/.

[noa, d] Java, http://www.oracle.com/technetwork/java/index.html.

[noa, e] Python, https://www.python.org/.

[noa, f] TreeAge, https://www.treeage.com/.

[haz, 1988] (1988). Turing Machine. In Hazewinkel, M., editor, *Encyclopaedia of mathematics*. Reidel ; Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers, Dordrecht ; Boston : Norwell, MA, U.S.A.

[Abler et al., 2013] Abler, D., Kanellopoulos, V., Davies, J., Dosanjh, M., Jena, R., Kirkby, N., and Peach, K. (2013). Data-driven Markov models and their application in the evaluation of adverse events in radiotherapy. *Journal of Radiation Research*, 54(suppl 1):i49–i55.

[Aoki, 2002] Aoki, M. (2002). *Modeling aggregate behavior and fluctuations in economics: stochastic views of interacting agents*. Cambridge University Press, New York.

[Austria, 2012] Austria, S. (2012). *Demographisches Jahrbuch Österreich 2012*. Verlag Österreich GmbH.

[Axelrod, 1997] Axelrod, R. (1997). *The complexity of cooperation: agent-based models of competition and collaboration*. Princeton studies in complexity. Univ. Press, Princeton, NJ. OCLC: 258486480.

[Axelrod and Tesfatsion, 2006] Axelrod, R. and Tesfatsion, L. (2006). Appendix A A Guide for Newcomers to Agent-Based Modeling in the Social Sciences. In *Handbook of Computational Economics*, volume 2, pages 1647–1659. Elsevier. DOI: 10.1016/S1574-0021(05)02044-7.

[Axelrod, 1984] Axelrod, R. M. (1984). *The evolution of cooperation*. Basic Books, New York.

[Baroni and Richiardi, 2007] Baroni, E. and Richiardi, M. (2007). *Orcutt's Vision, 50 years on*.

[Barthélemy and Carletti, 2017] Barthélemy, J. and Carletti, T. (2017). An adaptive agent-based approach to traffic simulation. *Transportation Research Procedia*, 25:1238–1248.

[Barton and Stamber, 2000] Barton, D. and Stamber, K. (2000). An agent-based microsimulation of critical infrastructure systems. Las Vegas, Nevada.

[Basu et al., 1998] Basu, N., Pryor, R., and Quint, T. (1998). ASPEN: A Microsimulation Model of the Economy. *Computational Economics*, 12(3):223–241.

[Belova and Murch, 2007] Belova, I. V. and Murch, G. E. (2007). Bridging Different Length and Time Scales in Diffusion Problems Using a Lattice Monte Carlo Method. *Solid State Phenomena*, 129:1–10.

[Bengtsson et al., 1984] Bengtsson, L., Simmons, A., Cats, G., Shukla, J., Leith, C., and Källèn, E. (1984). *Problems and Prospects in Long and Medium Range Weather Forecasting*. Springer Berlin Heidelberg, Berlin, Heidelberg. OCLC: 863816460.

[Benoit et al., 2006] Benoit, J., Nunes, A., and Telo da Gama, M. (2006). Pair approximation models for disease spread. *The European Physical Journal B*, 50(1-2):177–181.

[Berg et al., 1998] Berg, J. v. d., Geoffrey, R., Grimmett, R., and Schinazi, R. B. (1998). Dependent Random Graphs and Spatial Epidemics. *The Annals of Applied Probability*, 8(2):317–336.

[Bergero and Kofman, 2011] Bergero, F. and Kofman, E. (2011). PowerDEVS: a tool for hybrid system modeling and real-time simulation. *SIMULATION*, 87(1-2):113–132.

[Berlekamp, 2001] Berlekamp, E. R. (2001). *Winning ways for your mathematical plays*. A.K. Peters, Natick, Mass, 2nd ed edition.

[Bianchi and Squazzoni, 2015] Bianchi, F. and Squazzoni, F. (2015). Agent-based models in sociology: Agent-based models in sociology. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(4):284–306.

[Bicher, 2013] Bicher, M. (2013). *Agentenbasierte Modellbildung und Simulation auf Basis der Fokker-Planck-Gleichung*. Master Thesis, Inst. f. Analysis und Scientific Computing, Vienna University of Technology, Vienna, Austria.

[Bicher, 2016] Bicher, M. (2016). Proof for Mean-Field Theorems for Different State-Spaces.

[Bicher et al., 2015] Bicher, M., Glock, B., Miksch, F., Schneckenreither, G., and Popper, N. (2015). Definition, Validation and Comparison of Two Population Models for Austria. In *Proceedings of 4th UBT Annual International Conference on Business,Technology and Innovation*, Durres, Albania. UBT - Higher Education Institution.

[Bicher and Popper, 2013] Bicher, M. and Popper, N. (2013). Agent-Based Derivation of the SIR - Differential Equations. In Al-Begain, K., Al-Dabass, D., Orsoni, A., Cant, R., and Zobel, R., editors, *Proceedings of the 8th EUROSIM Congress on Modelling and Simulation*, pages 306–311, Cardiff, Wales.

[Bicher and Popper, 2015] Bicher, M. and Popper, N. (2015). Spatial Effects in Stochastic Microscopic Models - Case Study and Analysis. *IFAC-PapersOnLine*, 48(1):153–158.

[Bicher and Popper, 2016] Bicher, M. and Popper, N. (2016). Mean-Field Approximation of a Microscopic Population Model for Austria. In *Proceedings of the 9th EUROSIM Congress on Modelling and Simulation*, pages 544–545, Oulu, Finland.

[Bicher et al., 2017a] Bicher, M., Popper, N., and Schneckenreither, G. (2017a). Comparison of a microscopic and a macroscopic age-dependent SIR model. *Mathematical and Computer Modelling of Dynamical Systems*, 23(2):177–195.

[Bicher et al., 2017b] Bicher, M., Urach, C., Zauner, G., Rippinger, C., and Popper, N. (2017b). Calibration of a Stochastic Agent-Based Model for Re-Hospitalization Numbers of Psychatric Patients. In *Proceedings of the 2017 Winter Simulation Conference*. Accepted for Publication.

[Bolch, 1968] Bolch, B. W. (1968). The Teacher's Corner: More on Unbiased Estimation of the Standard Deviation. *The American Statistician*, 22(3):27–27.

[Bonabeau, 1997] Bonabeau, E. (1997). From Classical Models of Morphogenesis to Agent-Based Models of Pattern Formation. *Artificial Life*, 3(3):191–211.

[Bonabeau, 2002] Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, 99(Supplement 3):7280–7287.

[Breitenecker et al., 2013] Breitenecker, F., Bicher, M., and Wurzer, G. (2013). *Agent-based modeling and simulation in archaeology: Agent-Based Simulation in Archaeology: A Characterization*. Advances in Geographic Information Science. Springer, [S.l.].

[Breitenecker et al., 2004] Breitenecker, F., Popper, N., and Hötzendorfer, H. (2004). Temporal and Spatial Evolution of a SIR-type Epidemic - ARGESIM Comparison C17 - Definition. *Simulation News Europe SNE*, 15(41-42):42–44.

[Buss and Al Rowaei, 2010] Buss, A. and Al Rowaei, A. (2010). A comparison of the accuracy of discrete event and discrete time. pages 1468–1477. IEEE.

[Chen and Doolen, 1998] Chen, S. and Doolen, G. D. (1998). Lattice Boltzmann Method for Fluid Flows. *Annual Review of Fluid Mechanics*, 30(1):329–364.

[Corning, 2002] Corning, P. A. (2002). The re-emergence of emergence: A venerable concept in search of a theory. *Complexity*, 7(6):18–30.

[de Aguiar et al., 2003] de Aguiar, M., Rauch, E., and Bar-Yam, Y. (2003). Mean-field approximation to a spatial host-pathogen model. *Physical Review E*, 67(4).

[Deffuant et al., 2000] Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000). Mixing beliefs among interacting agents. *Advances in Complex Systems*, 03(01n04):87–98.

[Dembele,] Dembele, J.-M. Contribution to the modeling of complex physical systems described by PDE1: a 2d multiagent model for convectiondiffusion. *International Journal of Complex Systems*, (Article 2169).

[Dietz and Schenzle, 1985] Dietz, K. and Schenzle, D. (1985). Proportionate mixing models for age-dependent infection transmission. *Journal of Mathematical Biology*, 22(1).

[Durrett and Levin, 2000] Durrett, R. and Levin, S. (2000). Lessons on pattern formation from planet WATOR. *Journal of Theoretical Biology*, 205(2):201–214.

[Dynkin and Mandelbaum, 1983] Dynkin, E. B. and Mandelbaum, A. (1983). Symmetric Statistics, Poisson Point Processes, and Multiple Wiener Integrals. *The Annals of Statistics*, 11(3):739–745.

[Edwards et al., 2003] Edwards, M., Huet, S., Goreaud, F., and Deffuant, G. (2003). Comparing an individual-based model of behaviour diffusion with its mean field aggregate approximation. *Journal of Artificial Societies and Social Simulation*, 6(4).

[Ellner, 2001] Ellner, S. P. (2001). Pair Approximation for Lattice Models with Multiple Interaction Scales. *Journal of Theoretical Biology*, 210(4):435–447.

[Epstein, 2012] Epstein, J. M. (2012). *Generative Social Science*. Princeton University Press, Princeton. OCLC: 956983748.

[Farmer and Foley, 2009] Farmer, J. D. and Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256):685–686.

[Fibich and Gibori, 2010] Fibich, G. and Gibori, R. (2010). Aggregate Diffusion Dynamics in Agent-Based Models with a Spatial Structure. *Operations Research*, 58(5):1450–1468.

[Fone et al., 2003] Fone, D., Hollinghurst, S., Temple, M., Round, A., Lester, N., Weightman, A., Roberts, K., Coyle, E., Bevan, G., and Palmer, S. (2003). Systematic review of the use and value of computer simulation modelling in population health and health care delivery. *Journal of Public Health*, 25(4):325–335.

[Forrester, 1961] Forrester, J. W. (1961). *Industrial dynamics*. Productivity Press, Cambridge, MA, student edition edition.

[Frisch et al., 1986] Frisch, U., Hasslacher, B., and Pomeau, Y. (1986). Lattice-Gas Automata for the Navier-Stokes Equation. *Physical Review Letters*, 56(14):1505–1508.

[Gardiner, 2009] Gardiner, C. W. (2009). *Stochastic methods: a handbook for the natural and social sciences*. Springer series in synergetics. Springer, Berlin, 4th ed edition.

[Gast and Gaujal, 2011] Gast, N. and Gaujal, B. (2011). A mean field approach for optimization in discrete time. *Discrete Event Dynamic Systems*, 21(1):63–101.

[Gipps, 1981] Gipps, P. G. (1981). A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological*, 15(2):105–111.

[Glock et al., 2012] Glock, B., Einzinger, P., and Breitenecker, F. (2012). A modular architecture for modelling obesity in inhomogeneous populations in austria with system dynamics – first step: a populationmodel and how to integrate it in a diseasemodel. In *Proceedings of the European Modeling and Simulation Symposium, 2012*, pages 513–518, Vienna.

[Goldstein, 1999] Goldstein, J. (1999). Emergence as a Construct: History and Issues. *Emergence*, 1(1):49–72.

[Gordon, 1961] Gordon, G. (1961). A general purpose systems simulation program. In *Proceedings of the December 12-14, 1961, eastern joint computer conference: computers-key to total systems control*, pages 87–104. ACM.

[Gordon, 1978] Gordon, G. (1978). The development of the general purpose simulation system (GPSS). In *History of programming languages I*, pages 403–426. ACM.

[Grigoryev, 2012] Grigoryev, I. (2012). *AnyLogic 6 in three days: a quick course in simulation modeling*. AnyLogic North America, [Hampton, NJ].

[Hardy et al., 1976] Hardy, J., De Pazzis, O., and Pomeau, Y. (1976). Molecular dynamics of a classical lattice gas: Transport properties and time correlation functions. *Physical review A*, 13(5):1949.

[Hastings, 1970] Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97.

[Hey et al., 2009] Hey, T., Tansley, S., Tolle, K. M., and others (2009). *The fourth paradigm: data-intensive scientific discovery*, volume 1. Microsoft research Redmond, WA.

[Hoppensteadt, 1974] Hoppensteadt, F. (1974). An Age Dependent Epidemic Model. *Journal of the Franklin Institute*, 297(5):325–333.

[Ising, 1925] Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258.

[Jahn et al., 2016] Jahn, B., Conrads-Frank, A., Sroczynski, G., Rochau, U., Zauner, G., Bundo, M., Gyimesi, M., Endel, G., Popper, N., and Siebert, U. (2016). The Application of Population Models to Improve Access to Medical Innovations and Population Health – A Systematic Review. *Value in Health*, 19(7):A377.

[Jahn et al., 2014] Jahn, B., Rochau, U., Shterjovska, J., Kurzthaler, C., Kluibenschädl, M., Urach, C., Einzinger, P., Piringer, H., Popper, N., and Siebert, U. (2014). Visualizing methods for discrete-event-simulations using the example of a breast cancer decision-analytic model.

[Jaynes and Bretthorst, 2003] Jaynes, E. T. and Bretthorst, G. L. (2003). *Probability theory: the logic of science*. Cambridge University Press, Cambridge, UK ; New York, NY.

[Kadanoff, 2013] Kadanoff, L. (2013). Theories of Matter: Infinities and Renormalization. In *The Oxford handbook of philosophy of physics*, pages 141–188. Oxford University Press, New York.

[Kalman, 1960] Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45.

[Kampen, 2007] Kampen, N. G. v. (2007). *Stochastic Processes in Physics and Chemistry*. North-Holland personal library. Elsevier, Amsterdam ; Boston, 3rd ed edition.

[Kampen, N. G. van, 1982] Kampen, N. G. van (1982). The Diffusion Approximation for Markov Processes. In Lamprecht, I and Zotin, A.I., editors, *Thermodynamics and kinetics of biological processes*, pages 181–195. Walter de Gruyter & Co.

[Kermack and McKendrick, 1927] Kermack, W. O. and McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 115(772):700–721.

[Kim et al., 2009] Kim, J., Sosa, E., and Rosenkrantz, G. S., editors (2009). *A companion to metaphysics*. Blackwell companions to philosophy. Wiley-Blackwell, Malden, MA, 2nd ed edition.

[Kitzler, 2016] Kitzler, F. (2016). *Analyse von agentenbasierten Modellen mithilfe von Markovketten*. PhD thesis, TU Wien, Wien.

[Kolmogorov, 1933] Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, Heidelberg, New York. OCLC: 74144504.

[Kolmogorov, A, 1931] Kolmogorov, A (1931). Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen*, 104(1):415–458.

[Lachapelle and Wolfram, 2011] Lachapelle, A. and Wolfram, M.-T. (2011). On a mean field game approach modeling congestion and aversion in pedestrian crowds. *Transportation Research Part B: Methodological*, 45(10):1572–1589.

[Landau, 1937] Landau, L. D. (1937). On the Theory of Phase Transitions. *Zh. Eksp. Teor. Fiz.*, 7:19–32.

[Lander, 2015] Lander, W. (2015). *Agent-Based Models For Infectious Disease Transmission*. PhD thesis, University of Antwerpen, Antwerpen.

[Langton, 1990] Langton, C. G. (1990). Life at the edge of chaos. In *Artificial Life II: Proceedings of the Workshop on Artificial Life Held February, 1990 in Santa Fe, New Mexico (Santa Fe Institute Studies in the Sciences of Complexity Proceedings)*, pages 41–91. Addison-Wesley Publishing Company.

[Law, 2007] Law, A. M. (2007). *Simulation modeling and analysis*. McGraw-Hill series in industrial engineering and management science. McGraw-Hill, Boston, Mass., 4. ed., internat. ed edition. OCLC: 255955662.

[Levins, 1969] Levins, R. (1969). Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of the Entomological Society of America*, 15(3):237–240.

[Li and O'Donoghue, 2013] Li, J. and O'Donoghue, C. (2013). A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of Microsimulation*, 6(2):3–55.

[Li et al., 2016] Li, Y., Lawley, M. A., Siscovick, D. S., Zhang, D., and Pagán, J. A. (2016). Agent-Based Modeling of Chronic Diseases: A Narrative Review and Future Research Directions. *Preventing Chronic Disease*, 13.

[Li, 2007] Li, Y. C. (2007). On the True nature of turbulence. *The Mathematical Intelligencer*, 29(1):45–48.

[Ljubovic, 2009] Ljubovic, V. (2009). Traffic simulation using agent-based models. pages 1–6. IEEE.

[Lorenz, 1963] Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2):130–141.

[Lotka, 1956] Lotka, A. J. (1956). *Elements of mathematical biology.* Dover Publications, New York.

[Macal and North, 2006] Macal, C. M. and North, M. J. (2006). Tutorial on Agent-Based Modeling and Simulation Part 2: How to Model with Agents. In *Proceedings of the 2006 Winter Simulation Conference*, pages 73–83, Monterey, California.

[Matsuda et al., 1992] Matsuda, H., Ogita, N., Sasaki, A., and Sat\^{o}, K. (1992). Statistical mechanics of population: the lattice Lotka-Volterra model. *Progress of theoretical Physics*, 88(6):1035–1049.

[McKendrick, 1925] McKendrick, A. G. (1925). Applications of Mathematics to Medical Problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98.

[McMillan, 1965] McMillan, W. L. (1965). Ground State of Liquid He 4. *Physical Review*, 138(2A):A442–A451.

[Metropolis et al., 1953] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

[Miksch et al., 2015] Miksch, F., Bicher, M., Urach, C., Schneckenreiter, G., Glock, B., and Popper, N. (2015). Generic Population Handbook.

[Miksch et al., 2016] Miksch, F., Jahn, B., Siebert, U., Glock, B., Bicher, M., Schneckenreiter, G., Urach, C., and Popper, N. (2016). 2nd Asia-Pacific Conference of the Society for Medical Decision Making, An Approach to Comparative Population Modeling and Simulation. *Medical Decision Making*, 36(4):E1–E77.

[Miksch et al., 2011] Miksch, F., Urach, C., Popper, N., Zauner, G., Endel, G., Schiller-Frühwirth, I., and Breitenecker, F. (2011). New Insights on the Spread of Influenza Through Agent Based Modeling. Madrid, Spain.

[Moyal, 1949] Moyal, J. E. (1949). Stochastic Processes and Statistical Physics. *Journal of the Royal Statistical Society*, B(11):150–210.

[Murata, 1989] Murata, T. (1989). Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580.

[Nagel and Schreckenberg, 1992] Nagel, K. and Schreckenberg, M. (1992). A cellular automaton model for freeway traffic. *Journal de Physique I*, 2(12):2221–2229.

[Nance, 1996] Nance, R. E. (1996). A history of discrete event simulation programming languages. In *History of programming languages—II*, pages 369–427. ACM, New York, 1 edition. DOI: 10.1145/234286.1057822.

[Nianogo and Arah, 2015] Nianogo, R. A. and Arah, O. A. (2015). Agent-Based Modeling of Noncommunicable Diseases: A Systematic Review. *American Journal of Public Health*, 105(3):e20–e31.

[Nirei, 2006] Nirei, M. (2006). Threshold behavior and aggregate fluctuation. *Journal of Economic Theory*, 127(1):309–322.

[North, 2014] North, M. J. (2014). A theoretical formalism for analyzing agent-based models. *Complex Adaptive Systems Modeling*, 2(1):3.

[Orcutt, 1957] Orcutt, G. (1957). A new type of socio-economic system. *Review of Economics and Statistics*, 39(2):116–123.

[Parry, 2011] Parry, J. (2011). Visualisation for Decision-Makers. pages 8–8. IEEE.

[Railsback and Grimm, 2012] Railsback, S. F. and Grimm, V. (2012). *Agent-based and individual-based modeling: a practical introduction*. Princeton Univ. Press, Princeton. OCLC: 811181165.

[Raney et al., 2003] Raney, B., Cetin, N., Völlmy, A., Vrtic, M., Axhausen, K., and Nagel, K. (2003). An Agent-Based Microsimulation Model of Swiss Travel: First Results. *Networks and Spatial Economics*, 3(1):23–41.

[Reynolds, 1987] Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. pages 25–34. ACM Press.

[Rio, 2013] Rio, E. (2013). Inequalities and limit theorems for weakly dependent sequences.

[Rippinger, 2016] Rippinger, C. (2016). *Kalibrierung eines agentenbasierten Influenzamodells*. Master Thesis, TU Wien, Vienna, Austria.

[Rippinger et al., 2016] Rippinger, C., Bicher, M., and Miksch, F. (2016). Calibration Strategies for Agent-based Simulation Models with Variability. In Breitenecker, F., editor, *Tagungsband ASIM 2016 23. Symposium Simulationstechnik*, pages 239–245, Dresden, Germany. ASIM.

[Robinson, 2004] Robinson, S. (2004). *Simulation: the practice of model development and use*. Palgrave Macmillan.

[Rosenblatt, 1956] Rosenblatt, M. (1956). A Central Limit Theorem and a Strong Mixing Condition. *Proceedings of the National Academy of Sciences*, 42(1):43–47.

[Salgado et al., 2016] Salgado, D., Jolovic, D., Martin, P. T., and Aldrete, R. M. (2016). Traffic Microsimulation Models Assessment – A Case Study of International Land Port of Entry. *Procedia Computer Science*, 83:441–448.

[Satô et al., 1994] Satô, K., Matsuda, H., and Sasaki, A. (1994). Pathogen invasion and host extinction in lattice structured populations. *Journal of Mathematical Biology*, 32(3):251–268.

[Schelling, 2017] Schelling, T. (2017). Thomas C. Schelling - Biographical.

[Schelling, 1971] Schelling, T. C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, Vol. 1:143–186.

[Schelling, 1978] Schelling, T. C. (1978). *Micromotives and macrobehavior*. Fels lectures on public policy analysis. Norton, New York, 1st ed edition.

[Schneckenreither, 2014] Schneckenreither, G. (2014). *Developing Mathematical Formalisms for Cellular Automata in Modelling and Simulation*. Diploma Thesis, Vienna University of Technology.

[Schruben, 1983] Schruben, L. (1983). Simulation modeling with event graphs. *Communications of the ACM*, 26(11):957–963.

[Schruben, 1992] Schruben, L. W. (1992). SIGMA—A graphical approach to teaching simulation. *Journal of Computing in Higher Education*, 4(1):27–37.

[Scott, 1979] Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, 66(3):605–610.

[Shannon, 1958] Shannon, C. E. (1958). Von Neumann's contributions to automata theory. *Bulletin of the American Mathematical Society*, 64(3):123–129.

[Silverman, 1998] Silverman, B. W. (1998). *Density estimation for statistics and data analysis*. Number 26 in Monographs on statistics and applied probability. Chapman & Hall/CRC, Boca Raton.

[Skvortsov et al., 2007] Skvortsov, A., Connell, R., Dawson, P., and Gailis, R. (2007). Epidemic Modelling: Validation of Agent-based Simulation by Using Simple Mathematical Models.

[Spezzano and Talia, 1999] Spezzano, G. and Talia, D. (1999). Programming cellular automata algorithms on parallel computers. *Future Generation Computer Systems*, 16(2):203–216.

[Spielauer, 2007] Spielauer, M. (2007). Dynamic microsimulation of health care demand, health care finance and the economic impact of health behaviours: survey and review. *International Journal of Microsimulation*, 1(1):35–53.

[Stachowiak, 1973] Stachowiak, H. (1973). *Allgemeine Modelltheorie.* Springer-Verlag, Wien, New York.

[Stokes, 1845] Stokes, G. G. (1845). On the theories of internal friction of fluids in motion. *Trans. Camb. Philos. Soc.*, 8:287–305.

[Tainaka, 1989] Tainaka, K.-i. (1989). Stationary pattern of vortices or strings in biological systems: Lattice version of the Lotka-Volterra model. *Physical Review Letters*, 63(24):2688–2691.

[Tesfatsion, 2001] Tesfatsion, L. (2001). Introduction to the special issue on agent-based computational economics. *Journal of Economic Dynamics and Control*, 25(3-4):281–293.

[Tisue and Wilensky, 2004] Tisue, S. and Wilensky, U. (2004). NetLogo: A simple environment for modelling complexity. pages 16–21.

[Tocher and Owen, 2008] Tocher, K. D. and Owen, D. G. (2008). The automatic programming of simulations. *Journal of Simulation*, 2(3):143–152.

[Todhunter, 1865] Todhunter, I. (1865). *Histoy of Mathematical Theory of Probability*.

[Tokuzo and Hiroshi, 1985] Tokuzo, S. and Hiroshi, T. (1985). Central limit theorem for a system of Markovian particles with mean field interactions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 69(3):439–459.

[Treiber et al., 2000] Treiber, M., Hennecke, A., and Helbing, D. (2000). Congested traffic states in empirical observations and microscopic simulations. *Physical review E*, 62(2):1805.

[Ulam, 1951] Ulam, S. (1951). On the Monte Carlo Method. In *Proceedings of the 1949 Symposium on Large-Scale Digital Calculating Machines*, pages 207–229, Cambridge, Massachusetts. Harward University Press.

[Ulam, 1982] Ulam, S. (1982). Biography of John von Neumann. *IEEE Annals of the History of Computing*, 4(2):157–181.

[Uzochukwu et al., 2016] Uzochukwu, B., Onwujekwe, O., Mbachu, C., Okwuosa, C., Etiaba, E., Nyström, M. E., and Gilson, L. (2016). The challenge of bridging the gap between researchers and policy makers: experiences of a Health Policy Research Group in engaging policy makers to support evidence informed policy making in Nigeria. *Globalization and Health*, 12(1).

[Van Dinther, 2007] Van Dinther, C. (2007). Agent-based Computational Economics. In *Adaptive Bidding in Single-Sided Auctions Under Uncertainty: An Agent-based Approach in Market Engineering*, pages 47–68. Birkhäuser Basel, Basel. DOI: 10.1007/978-3-7643-8113-4_3.

[Verdon-Kidd et al., 2014] Verdon-Kidd, D. C., Kiem, A. S., and Austin, E. K. (2014). Bridging the gap between researchers and decision-makers. In Palutikof, J. P., Boulter, S. L., Barnett, J., and Rissik, D., editors, *Applied Studies in Climate Adaptation*, pages 51–59. John Wiley & Sons, Ltd, Chichester, UK. DOI: 10.1002/9781118845028.ch7.

[von Winterfeldt, 2013] von Winterfeldt, D. (2013). Bridging the gap between science and decision making. *Proceedings of the National Academy of Sciences*, 110(Supplement_3):14055–14061.

[Waldherr and Wijermans, 2013] Waldherr, A. and Wijermans, N. (2013). Communicating Social Simulation Models to Sceptical Minds. *Journal of Artificial Societies and Social Simulation*, 16(4).

[Warnke et al., 2015] Warnke, T., Steiniger, A., Uhrmacher, A. M., Klabunde, A., and Willekens, F. (2015). ML3: A language for compact modeling of linked lives in computational demography. pages 2764–2775. IEEE.

[Webb et al., 2007] Webb, S. D., Keeling, M. J., and Boots, M. (2007). Host–parasite interactions between the local and the mean-field: How and when does spatial population structure matter? *Journal of Theoretical Biology*, 249(1):140–152.

[Wilensky,] Wilensky, U. NetLogo, https://ccl.northwestern.edu/netlogo/.

[Winkler, 2014] Winkler, S. (2014). *Comparative mathematical modelling of groundwater pollution*. Diploma Thesis, TU Wien, Wien, Austria.

[Winkler et al., 2015] Winkler, S., Bicher, M., and Breitenecker, F. (2015). Alternative approaches for groundwater pollution. *IFAC-PapersOnLine*, 48(1):159–164.

[Wolfram, 1984] Wolfram, S. (1984). Universality and complexity in cellular automata. *Physica D: Nonlinear Phenomena*, 10(1-2):1–35.

[Wolfram, 2002] Wolfram, S. (2002). *A new kind of science*. Wolfram Media, Champaign, IL.

[Wood, 2012] Wood, S. (2012). Traffic microsimulation–dispelling the myths. *Traffic Engineering and Control*, 53(9):339–344.

[Young and Elcock, 1966] Young, W. M. and Elcock, E. W. (1966). Monte Carlo studies of vacancy migration in binary ordered alloys: I. *Proceedings of the Physical Society*, 89(3):735–746.

[Zauner et al., 2017] Zauner, G., Urach, C., Bicher, M., Popper, N., and Endel, F. (2017). Spatial psychiatric hospitalization modelling in an international setting - an agent based approach. In *Proceedings of the International Workshop on Innovative Simulation for Health Care 2017*, Barcelona, Spain. To Appear.

[Zechmeister, 2017] Zechmeister, M. (2017). Vergleich von zeitgetakteter und eventgesteuerter agentenbasierter Modellierung anhand eines Bevölkerungsmodells von Österreich.

[Zeigler, 1976] Zeigler, B. P. (1976). *Theory of Modelling and Simulation*. John Wiley & Sons, Inc., Canada.

About the Author

Martin Bicher (ORCID iD 0000-0002-1362-6868) studied Technical Mathematics at the Technical University Vienna (TU Wien). His interests led him to specializing in Modelling and Simulation with focus on microscopic modelling. In his Diploma Thesis 'Agent-based Modelling and Simulation on Basis of the Fokker-Planck - Equation' (in German) he already dealt with the upcoming demand for a solid foundation of microscopic modelling. This experience offered him a position as a researcher at TU Vienna and at dwh Simulation Services, where he started to work on population dynamic models, in conjunction with comparative modelling techniques on macroscopic, microscopic, and mesoscopic approaches. Within several projects he developed the mean field theory as common denominator for microscopic and macroscopic modelling, in theory and in application – resulting in this thesis and in several scientific publications. As the head of the COVID-19 simulation model development team of dwh GmbH, he currently applies the theory for analysis of epidemiological models. Hereby he and his team play a leading role in counselling the Austrian Ministry of Health in the time of the 2020/21 SARS-CoV-2 pandemic.

About the Book

Agent-based modelling and simulation versus ODE/PDE modelling – an unbridgeable antagonism in modelling and simulation? This thesis proofs existence and efficient use of a bridge – the mean field theory, in theory and application. The author presents a very well written mixture of monograph and research work summary. Using a valid mean-field model, i.e. a macroscopic equation-based model that approximates microscopic model results, the author utilizes the excellent analysis features of equation based approaches to analyse the microscopic model. He uses case studies from population dynamics, epidemiology, physics (surprising: an agent-based model for the pendulum), and game theory. And additionally he shows how mean-field models can be applied directly for sensitivity analysis and parameter calibration. As final result, the author proposes a new classification concept for microscopic models: a series of attributive adjectives according to the model's time-update, state-space, randomness and interaction do not only convey a unique picture of specific parts of the model, but also give ideas on possible challenges involved with model, simulation, parametrisation, sensitivity, and finally its mean-field behaviour.

About the Series

The ASIM series Advances in Simulation / Fortschrittsberichte Simulation presents new and recent approaches, methods, and applications in modelling and simulation. The topics may range from theory and foundations via simulation techniques and simulation concepts to applications. As the spectrum of simulation techniques and applications is increasing, books in these series present classical techniques and applications in engineering, natural sciences, biology, physiology, production and logistics, and business administration, upcoming simulation applications in social sciences, media, data management, networking, and complex systems, and upcoming new simulation techniques as agent-based simulation, co-simulation, and deep learning, etc.

The series puts emphasis on monographs with special character, as PhD theses, habilitation treatises, project reports and overviews on scientific projects. ASIM - Arbeitsgemeinschaft Simulation, the German Simulation Society (part of GI – Gesellschaft für Informatik) has founded the series Advances in Simulation / Fortschrittsberichte Simulation together with ARGESIM Publisher Vienna in order to provide to the international simulation community a quick and cost-efficient print and e-book series with open access.

ISBN ebook
ISBN 978-3-903347-29-8

DOI ID
DOI 10.11128/fbs.29

ARGESIM Publisher, Vienna, 2020
www.argesim.org