

Diploma Thesis
Diplomarbeit im Fach Physik*

Panic Reactions and Global Disease Dynamics

A spatially continuous mean-field approach without milk

vorgelegt vom Christian Thiemann
aus Osnabrück

angefertigt 2008
im Institut für Nichtlineare Dynamik
der Georg-August-Universität Göttingen



* Kann Spuren von anderen Wissenschaften enthalten.

Diplomarbeit im Fach Physik

Panic Reactions and Global Disease Dynamics

A spatially continuous mean-field approach without milk

Christian Thiemann

Betreuer: Dr. Dirk Brockmann
Erstprüfer: Prof. Dr. Theo Geisel
Zweitprüfer: Prof. Dr. Reiner Kree

Design Principles

This thesis is not a scientific paper. I deliberately chose not to write in a very formal style without any adjectives and such. Instead, this is more like a textbook. Of course, it is not really a textbook, but it is rather a textbook than a paper. I used the phrases “that is” and “i.e.” a lot, although you often hear that one should avoid such constructs because they reveal that the text before the phrase is so complicated that it does not explain anything while the following text contains all the information. I object to that and use them when two different formulations are available. According to my experience, reading things twice in different formulations clarifies things much better than reading the same thing over and over again. Furthermore, I absolutely disagree with the advocates of “avoid ‘we,’ never use ‘I,’ and use passive voice instead.” The readability of texts has been found to be by nothing else more improved than by this rule being disregarded, i.e., I found that nothing else improves readability more than disregarding this rule.

I also intentionally left out any generic theoretical background information to prevent the text from being cluttered up with too many side stories. I wanted to directly lead to the results, giving every little detail that is necessary but nothing more. However, where explanations of general concepts seemed to be useful or necessary to give a more comprehensive picture, they were added as a footnote. The text is also not intended as a presentation of results to a group of specialists but aims at a broader audience.

You may have noticed that the thesis is laid out on landscape paper. There are two main reasons for this: First, reading documents in electronic form becomes more and more popular while at the same time computer displays tend to become wider. With the traditional portrait orientation, pages have to be displayed piecewise or cover only half of the screen area when viewed in full. In contrast, the page of a landscape document neatly fits today’s display dimensions. Second, typographers consider a line length of at most 60 to 70 characters to be optimally readable. Compliance with this rule normally requires rather large page margins or typesetting in two columns. However, in portrait orientation, the columns are necessarily narrow and long—a format that significantly deviates from the aesthetic proportions of the paper itself (which are somewhat close to the Golden

Ratio). The landscape document allows for small page margins, and thus optimal resource usage, by dividing the page into two columns, each of which provides an aesthetically pleasing text area that also observes the line length rule.

Following the idea of electronic readability, I chose Concrete Roman as the main font for the document. Some people advise to use sans-serif fonts for digitally presented text because the thin serifs tend to be poorly rendered on low-resolution displays. However, serif fonts have a long tradition in printing because the serifs help to visually group letters and especially in guiding the eye along the line of text. The Concrete font created by Donald E. Knuth has a strong advantage: it uses a nearly constant stroke width, while in the standard L^AT_EX font, Computer Modern, and the popular Times, letters are partly composed of very thin lines. Thus, on a computer display, pages typeset in Concrete are readable even when scaled down to small sizes because there are no thin lines which disappear or are smeared out in font anti-aliasing *and* there are serifs to guide the reader’s eye. Last but certainly not least, Concrete’s slightly typewriter-ish appearance tints the document with a beautiful, old-fashioned look. The font chosen for the chapter and section titles is the sans-serif Avant Garde, which is very contrasty to the old-fashioned style of the main font while not looking misplaced. Unfortunately, it is too contrasty to be used with Concrete in the same line and there is no bold face of the latter, so I had to use Computer Modern Bold for the title part of the figure captions.

Speaking of figures: When preparing a paper, the included figures are often required to be black-and-white compatible. However, color as an independent information dimension allows producing very compact figures of high information density. In this thesis, I deliberately ignored b&w-compatibility, though mainly for aesthetic reasons: it adds unnecessary visual complexity like different line styles and often requires use of less pleasing color palettes.

One last remark: For best reading experience, the reader should assume that every unusual use of words or phrases and unannounced non-compliance with typographic conventions or those regarding “proper scientific style” is intended and meant as an expression of sheer creativity or just as a joke. In most instances this is indeed the case.

Contents

1	Introduction	1
1.1	Synopsis of Theory and Methods	1
2	Theory	7
2.1	Society of Classes: Stochastic Compartmental Models in Epidemiology	7
2.2	Ignoring Chances: The Mean-field Approximation	9
2.3	Megacities and Ghost Towns: Variable Population Density	13
2.4	Population Networking: Spatially Extended Systems	14
2.5	Indiscrete Models: From Compartments to Continuous Variables	15
2.5.1	Travelling Waves in Spatially Continuous Systems	18
2.6	Populations that Realize: Different Response Scenarios	20
2.6.1	Panic Reaction: Additional Diffusion	20
2.6.2	Directed Flight: Gradient Force	20
2.6.3	Strategic Flight: Integral Force	20
2.6.4	Corresponding Strategies in Stochastic Models	21
3	Methods	23
3.1	Concepts	23
3.1.1	Solving Ordinary Differential Equations	23
3.1.2	Solving Partial Differential Equations	24
3.1.3	Following Travelling Waves with a Moving Simulation Window	26
3.2	Details	26
3.3	Details' Details	28
4	Results	33
5	Discussion	39
5.1	Conclusions	39
5.2	Relevance	40
5.3	Outlook	41

1 Introduction

Epidemiological research in general most probably needs no justification. Its fundamental objective, to understand the mechanisms of human diseases and possible intervention strategies to prevent outbreaks or at least lessen the consequences thereof, is easily recognized as being important. Classical epidemiological research in the medical sciences focuses on capturing the current state of a population with respect to a certain disease (e.g., the prevalence of the disease), statistical correlations among risk factors from which causalities like “smoking increases the cancer risk” might be inferred, and small-scale dynamics like “who infected whom?” by back-tracing contacts of infected individuals. Aside from this, an increasing group of scientists is trying to understand the dynamics of diseases on a global scale.

Early work from 1927 [Ker27] provided mathematical models to describe the dynamics of a disease (i.e., the number of infected individuals over time) in a single population like a school or a small village. More sophisticated models were developed to take into account the interaction between several villages or town districts, providing descriptions of the dynamics in a country of the Middle Ages. The basic assumption behind those models is that people can travel to neighboring villages only. Modern technology, however, has led to a dramatic change of the human travel behavior. Today it is possible to cross the whole country in one day by car or train—or even across continents by plane. While ancient diseases like the Black Death spread over the country in a slow wave, modern diseases like SARS spread over the whole world within days. Thus, while ancient diseases can be modeled by agents on a regular grid, for modern diseases we have to take into account the complex travel networks that humans form in their everyday life.

Furthermore, early models used to use homogeneous populations, i.e., they assumed that there is a “typical personality” which can be used to represent every person of the population. In the SARS pandemic, however, very few people infected many more people than most of the other infecteds did. These so-called superspreaders play an important role and for many diseases, especially for sexually transmitted ones, the distribution of contact rates of individuals obey a power law, i.e., there is no “typical” contact rate valid for all individuals of the population. Thus, the social heterogeneity of a population can have significant impact on the

disease dynamics and has to be captured by accurate models.

Modern epidemiological research in the mathematical and complex dynamics departments therefore focuses on the understanding of complex social and geographical heterogeneities and their impact on global diseases. With better understanding of these factors, more effective actions can be proposed to prevent global outbreaks of emerging diseases or work towards the extinction of endemic diseases.

In this study, I focus on the impact of the geographical aspects of a population’s response to a disease outbreak. In particular, I will use spatially continuous reaction-diffusion models of socially homogeneous populations as a basic framework and define three kinds of response scenarios: A random panic reaction, a more intelligent directed flight reaction, and a strategic flight reaction. Though the underlying framework does not correctly capture modern human travel behavior, it still provides an effective testbed for investigating the basic mechanisms in the scenarios. In these models, as mentioned above, diseases spread via travelling waves. Qualitatively, the response may accelerate or slow down the disease spread, which can be quantized by measuring the equilibrium wave front velocity. Furthermore, the population’s reaction might lead to outbreak prevention, thus I will also investigate the initial non-equilibrium dynamics.

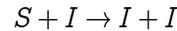
While most of the theory in Chapter 2 can be found in textbooks on mathematical chemistry or biology that cover reaction-diffusion systems [Mur89, Gar85], Sec. 2.6, which introduces the systems with response scenarios, presents, to my knowledge, original work—as do, of course, the results and discussion chapters. The methods presented in Chapter 3 can also be found in any textbook on computational science [Pre02], but the systematic investigation of the accuracy in Sec. 3.3 again presents own work, though parts of it probably have been published or at least investigated earlier already.

1.1 Synopsis of Theory and Methods

In this section I will give a figurative and intuitive presentation of the theoretical background and utilized methods. This should make it possible for the experienced

and impatient reader, who is not interested in the details of the underlying theory and methods, to directly jump to the Results and Discussion chapters starting on page 33. But primarily, this chapter aims at making it possible for interested people without strong physical background to do the same.

Infectious diseases can be modeled as chemical reactions. It's that simple: Assign a label S to healthy, susceptible people, and a label I to infected people. Then, an infected person can infect a susceptible, which is expressed by the following chemical reaction:



Furthermore, the reaction $I \rightarrow S$ expresses that an infected person can spontaneously, i.e., without having to meet somebody else, become healthy again. This simple model of the life cycle of a sick person is called the *SIS model*.

We can replace the reaction with $I \rightarrow R$ and have defined the *SIR model*. Here, we have another class of people labelled R , which are recovered. Since there is no reaction converting recovered people into infecteds, this essentially means that people gain immunity when they become healthy.

An interesting quantity is the number of infected people in a population which I will denote with I . Of course, this number changes over time and thus, we are interested in the dynamics of the disease, i.e., the time evolution of I . Within a certain time interval Δt , say, one week, a certain number of individuals will recover from a disease. Let's assume that infected people stay infected for about two weeks. Let's also assume that we have a bunch of infected people that have been infected at various times in the past. Then, after two weeks, all those people will be healthy again. Since we do not know the exact infection time of every individual we would expect that after one week about half of all the people will be healthy. To generalize from the disease with a duration of two weeks to arbitrary diseases, we introduce a model parameter β that states the fraction of people recovering from the disease after one week, or more generally, within the time interval Δt . The total number of infected people recovering within this time interval Δt is then given by $\beta I \Delta t$.

A bit more tedious to derive is the positive contribution to the number of infecteds per time unit, i.e., the number of infection events. For this to happen, an infected and a susceptible person have to meet. We assume that everybody in the population can meet everybody else with equal probability (this is not true in reality—but it simplifies the model). Then, if an infected person meets somebody else, the probability of this other person being a susceptible person is S/N , the fraction of susceptible persons in the population. We introduce another model parameter α that states the number of contacts a person has per time unit. In fact, α

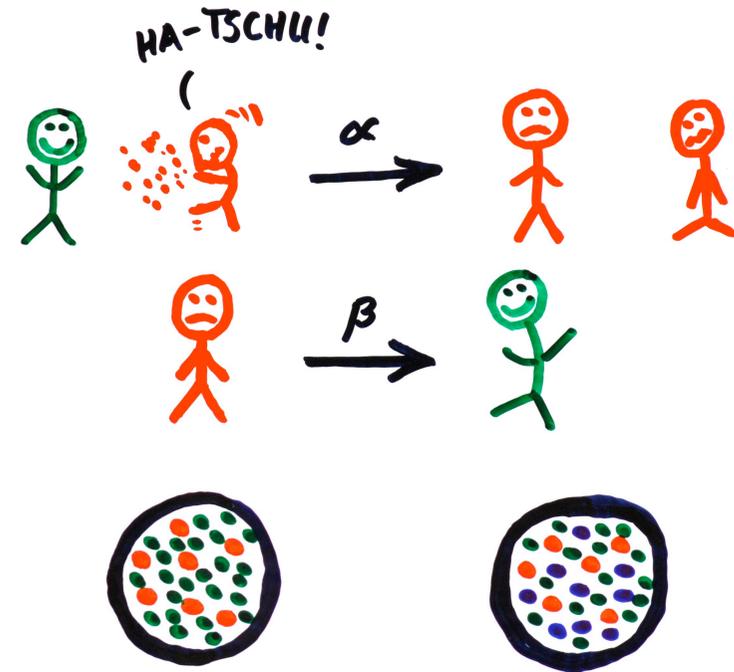


Figure 1.1: The SIS and SIR models. Essential ingredient of both models is the assumption of a well-mixed population (circles at the bottom) where everybody can meet everybody else with equal probability. In the SIS model, there are only two types of individuals in the population: susceptible (healthy) people and infected people (cf. left circle). The SIR model adds a third type, the recovered people (right circle). The “chemical” reactions shown in the top are those of the SIS model. The only difference in the SIR model is that in the recovery reaction (the bottom one) an infected does not become a susceptible (green) but a recovered person (blue) that cannot be infected again.

does not really count *every* contact but only those contacts in which the infected person did something that transmits the disease to the other person (i.e., sneezing). Then, in the time interval Δt , one infected person will meet $\alpha \Delta t$ persons and transmit the disease. But only in those cases where a susceptible person is met, a new person is actually infected. Thus, every infected person infects $\alpha \Delta t S/N$ persons. And then, since there are I infected persons in the population, a total number of $\alpha \Delta t I S/N$ persons get infected every week. One small modification is

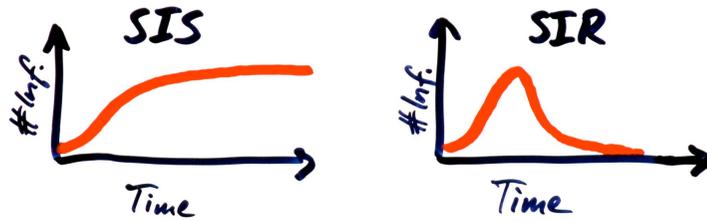


Figure 1.2: Qualitative mean-field behavior of the SIS and SIR model. In the SIS model, the number of infected people will reach a stable non-zero level, the so-called *endemic state*, since the recovery reaction $I \rightarrow S$ ensures an everlasting “supply” for new infections. In the SIR model, the reaction reads $I \rightarrow R$ and thus no continuous supply of susceptibles exist. Also, the above behavior is only observed if the infection rate α is larger than the recovery rate β . If $\alpha < \beta$, the disease will quickly die out, because infected persons will, on average, recover faster than they infect new people.

to be done when taking into account the total density of people. Obviously, the probability of meeting somebody is larger in dense populations while it is smaller in sparsely populated areas. Therefore, we expect the number of infectious contacts to be proportional to the density $\rho = N/\Omega$ of the N people living in an area Ω .

Put together, in the time interval Δt the number of infected people I changes by

$$\Delta I = \alpha \rho \frac{SI}{N} \Delta t - \beta I \Delta t.$$

We can do serious math tricks here by making Δt very small and obtain the differential equation

$$\frac{dI}{dt} = \alpha \frac{SI}{\Omega} - \beta I.$$

Furthermore, we will rewrite our model in terms of the density of infecteds, $j = I/\Omega$, and density of susceptibles, $s = S/\Omega$:

$$\frac{dj}{dt} = \alpha s j - \beta j.$$

The differential equations for dj/dt are the same in the SIS and the SIR model, but while $ds/dt = -dj/dt$ in the SIS model, we have an additional variable in the SIR model, namely the density of recovered people, $r = R/\Omega$. Here, the term $+\beta j$ does not appear in the equation for ds/dt but in the one for dr/dt , since in

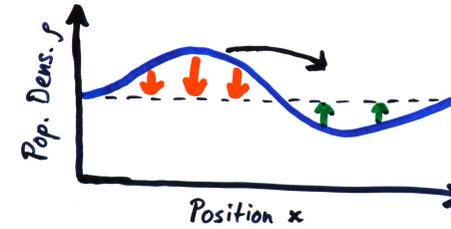


Figure 1.3: Diffusion. The figure shows a population density profile (blue line) and indicates where the diffusion term leads to a local increase of the density (green arrows) or to a decrease (red). Effectively, this can be interpreted as people moving from dense to sparsely populated areas (black arrow).

the SIR model recovered people are not susceptible again but belong to their own class. The dynamics of the SIS and the SIR model are depicted in Fig. 1.2.

We can easily add a geographical dimension to the model by redefining the densities as functions of space and time: $j(x, t)$, $s(x, t)$, and $r(x, t)$, where at each location the total density is $\rho(x, t) = s(x, t) + j(x, t) + r(x, t)$ (with $r(x, t) = 0$ in the SIS model). Now, the above differential equation describes the *local* dynamics of the disease at each location x . Adding that people are slowly and randomly moving around, the density of infecteds obeys the following equation, which is enriched by a *diffusion term*:

$$\frac{\partial j}{\partial t} = \alpha s j - \beta j + \frac{\partial^2 (D j)}{\partial x^2}.$$

The *diffusion coefficient* D is also a function of the location, $D = D(x, t)$. Of course, $\partial^2 (D s)/\partial x^2$ and $\partial^2 (D r)/\partial x^2$ are added to the respective equations. What is the effect of this? Since the second derivative indicates curvature, the diffusion term is positive for convex parts of the function and negative for concave parts (cf. Fig. 1.3). For constant diffusion coefficients $D(x, t) = D_0$, this leads to a leveling of the population density, i.e., after some large time t , the population density profile $\rho(x, t)$ will essentially look flat, and for $t \rightarrow \infty$, it is constant: $\rho(x, t) = \rho_0$.

The diffusion terms are not effective on the total density ρ , but on the densities s , j , and r . This means that for $t \rightarrow \infty$, each individual density will level out to constant functions (and thus, the previous claims regarding the total population density follow). This also means that local deviations from these desired flatness are immediately “smeared out.” In particular, a disease emerging at a certain location

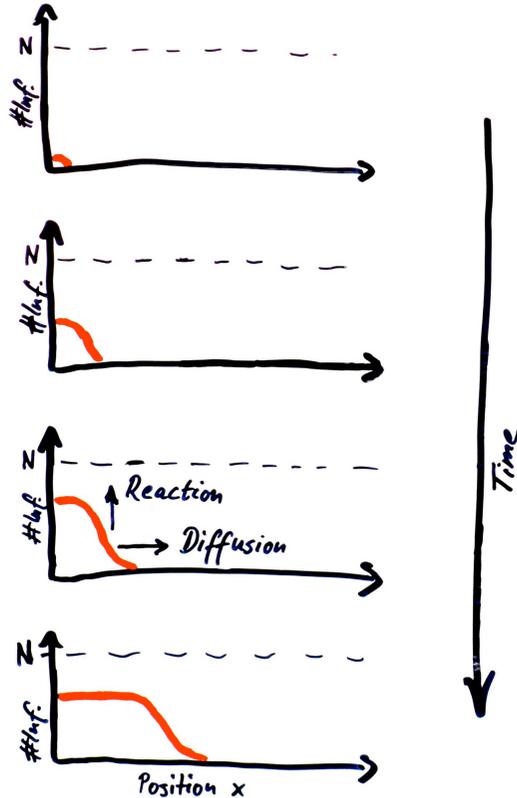


Figure 1.4: Emerging travelling waves in spatially extended systems. The figure shows a series of snapshots of the number of infecteds or density of infecteds in a one-dimensional space. Initially, a small number of infected people is placed at $x = 0$ and, successively, the local reactions produce more infected people at that location while the diffusion carries the infected people towards neighboring places. The interplay of both effects eventually leads to a stable wave front that moves across the system.

x_0 , indicated by a raise of the density $j(x_0, t)$, will by diffusion quickly lead to an increasing density of the neighboring sites, $j(x, t)$, $x \approx x_0$. The combination of local disease dynamics (reactions), which, as long as j is small and s is large, exhibit exponential growth of j , and the diffusion process leads to the emergence of epidemic waves rolling over the population (cf. Fig. 1.4).

As the interested reader probably quickly figures out, the diffusion process is undirected, though its strength can be locally different. I would also like to model population responses where people run away from an emerging disease, thus where the travel behavior is *directed*. We can achieve this by adding a *drift term* to all the density equations:

$$\frac{\partial j}{\partial t} = \alpha s j - \beta j + \frac{\partial^2(D j)}{\partial x^2} - \frac{\partial(F j)}{\partial x}.$$

Again, the *drift coefficient* F is a function of space and time, $F = F(x, t)$. The effect of the drift term is that people at location x feel a force to travel to the right if $F(x, t) > 0$ or to the left if $F(x, t) < 0$. Thus, contrary to the undirected diffusion, people prefer to travel in one direction.

In this framework we can motivate three different kinds of responses to an epidemic wave running through the system (Fig. 1.5). The first scenario defines a diffusion coefficient that is proportional to the local density of infecteds (the parameter ν controls the strength of the response),

$$D(x, t) = D_0 (1 + \nu j(x, t)), \quad F(x, t) = 0.$$

This can be interpreted as a simple *panic reaction* where people in infected areas feel the urge to move away, no matter in what direction.

The second scenario models a *directed flight* where people compare the level of infection at their own location with that in their immediate environment and move towards regions with lower density of infecteds. This is expressed by the spatial derivative of density of infecteds in the drift coefficient (again, μ is a control parameter indicating the response strength),

$$D(x, t) = D_0, \quad F(x, t) = -\mu \frac{\partial j}{\partial x}.$$

In the third, *strategic flight* scenario, the previous ansatz is extended to an integral force

$$F(x) = -\mu \int_{-\infty}^{\infty} \frac{x-y}{|x-y|} K(|x-y|) j(y) dy.$$

This is a generalized gravitational force where the density of infecteds is to be interpreted as *repelling* mass. The integration kernel $K(d)$ determines the influence of infecteds at a distance d on the repelling effect. In this study I use the kernel

$$K(d) = Z \left(\exp \left(-\frac{(d - d_{\max})^2}{2\sigma^2} \right) + \exp \left(-\frac{(d + d_{\max})^2}{2\sigma^2} \right) \right),$$

which, depending on the parameters d_{\max} and σ is rather small for $d = 0$, raises to a peak at $d = d_{\max}$, and then falls back to zero for $d > d_{\max}$. This means that infecteds which are far away have no repelling effect (they are not perceived), infecteds at a distance $d \approx d_{\max}$ have a strong repelling effect (they are perceived as a risk factor), and infecteds at a close distance $d \gtrsim 0$ have, if at all, a weak repelling effect (they are perceived as a risk factor but a flight is considered either useless, since one will be reached by the wave anyway, or unethical, since one's own flight might further facilitate spread of the disease into currently unaffected regions).

In this work, I study the effect of the three response scenarios on both the speed and shape of the travelling wave front. The differential equations presented above are discretized into a one-dimensional lattice with distance Δx between the nodes and integrated in time using the Runge-Kutta method. To provide a sufficiently large simulation area with low computational effort, a shifting window mechanism is used: The system is initialized with a small fraction of infecteds at the left end of the simulation area (as in Fig. 1.4) and each time the wave reaches a certain point of the simulation area, one lattice site is removed from the left and one is added to the right. This way, I follow the wave in a moving frame which allows me to observe it for long times. The wave front speeds and shapes are measured after some equilibration time ensuring that a fully developed wave front has emerged. Non-equilibrium dynamics, i.e., initial outbreak behaviors, are also investigated in a two-dimensional system, which naturally rises from the one-dimensional equation introduced above ($\nabla = (\partial/\partial x, \partial/\partial y)^T$ denotes the two-dimensional spatial derivative):

$$\frac{\partial j}{\partial t} = \alpha s j - \beta j + \nabla^2(D j) - \nabla(F j).$$

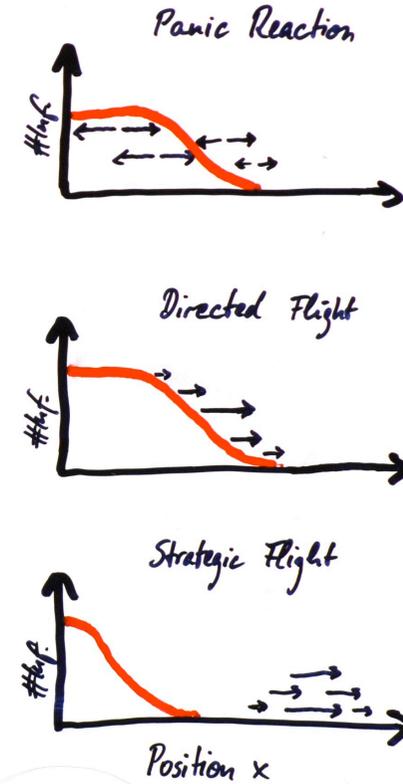


Figure 1.5: Response scenarios investigated in this study. Above: The plots show the number of infecteds (or density of infecteds) in a one-dimensional space. The black arrows visualize the effect of the different response scenarios by indicating the effective motion of people. Below: The integration kernel used in the strategic flight scenario.



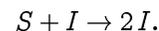
2 Theory

In this chapter, I will present the theoretical background of this study. I start with introducing the long used compartmental models, which describe the disease dynamics in a single well-mixed population, and point out the most important aspects of their dynamics. Averaging out stochastic fluctuations in large population sizes yields the deterministic mean-field approximation of the models, as presented in Sec. 2.2. Next, the compartmental models are reformulated to incorporate population density, necessary to reflect local density fluctuations in spatially extended systems which are then presented in Sec. 2.4.

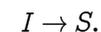
Section 2.5 introduces geographical structure to the models by embedding a linear chain of cities into a one-dimensional space. By going to the continuum limit we obtain a spatially continuous model which serves as the foundation of the study. In the framework of this model, I will finally introduce three different response strategies to a forthcoming epidemic wave (Sec. 2.6).

2.1 Society of Classes: Stochastic Compartmental Models in Epidemiology

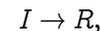
Compartmental models as used in epidemiological modeling since over eight decades [Ker27] are derived from chemical reaction models. The idea is simple: We have a population of, say, N individuals. We now assume that this population is *well-mixed*, that is, every individual has a chance to meet every other individual in the population and these chances are equal for all possible contacts. Now comes chemistry: We classify every individual in the population as a certain type and then define reactions between different types of individuals. In epidemiology, the interaction between healthy and sick people is investigated, so we introduce two classes: *Susceptibles*, which are susceptible to infection, and *infecteds*, which carry an infectious disease and are able to pass it on to a susceptible person. This passing-on is modeled as a chemical reaction of the form



Thus, we have an auto-catalytic reaction where infected people transform susceptibles into infecteds. Of course, infected people will become healthy after some time, which we can model with a spontaneous reaction like



This model is known as the *SIS model* since every individual will run through cycles of being susceptible, infected, susceptible etc., and can be used to model diseases like the flu. However, there are other diseases where infected people can become healthy again but are then immune to the disease (pox, for example). This can be modeled by introducing a new class of individuals, the *recovereds* or *removeds*. In fact, there are diseases where people stop being infectious because they actually die rather than becoming healthy—from the modeling point of view, we don't care about the difference. If we modify the spontaneous reaction $I \rightarrow S$ to read



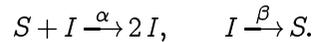
we captured exactly what happens to pox-infected people: After some time they transit from the infected state into a recovered state without being susceptible again, i.e., without being able to participate in the auto-catalytic $S + I \rightarrow 2I$ reaction. Analogously to the SIS naming convention, this model is then called the *SIR model*.

These two, SIS and SIR, are the most fundamental models in epidemiological modeling. There is an extension of SIR, called the *SIRS model*, which adds a third reaction of the form $R \rightarrow S$, i.e., recovered people can become susceptible again. This is a more accurate model of, for example, the flu: one actually is immune to the particular strain of flu he or she was infected with and only after some time the flu will have evolved into a new form to which this person is then susceptible again. Speaking of time, this is another crucial ingredient to the chemical reaction models. But let me first shortly introduce another family of compartmental models popular in epidemiology: We can define another class of individuals, the *exposed* people, which represent persons already carrying the disease without being able to

pass it on. This is an even more accurate model of the flu, since it can account for the incubation time, but in our area of interest it does not qualitatively change the overall dynamics. However, just to be comprehensive, there are *SEIS*, *SEIR*, and *SEIRS* models and their names already describe how they work. To give one example, the SEIR model is described by this set of reactions:



Back to the time issue. Every reaction occurs at a certain rate which is noted above the reaction arrow. Thus the complete SIS model is given by



The reaction rate states how likely the reaction occurs within a given time period and under the condition that all participating individuals (on the left hand side) can easily meet. We will see what this rate exactly means when going on to expressing the chemical reactions in mathematical equations.

Let us model our SIS population as a tuple of two numbers, (S, I) , stating the number of susceptibles and infecteds, respectively. Of course, nobody can hide from our classification system, thus $S + I = N$ (in an SIR model we would have a triple (S, I, R) with $S + I + R = N$, but let us focus on SIS for a moment). Note the simplification that underlies this model: We are not interested in the health state of every single individual person, i.e., we do not track whether John Smith is sick or healthy, but we are only interested in the total number of infected and susceptible people. Now, given (S, I) at a time t , what is the value (S, I) at a short time Δt later? Obviously, as S decreases, I will increase, and vice versa, due to the conservation law $S + I = N$.

How can S be decreased? The only way to make this happen is by an infection reaction which transforms a susceptible into an infected person. Thus, the interesting question is how probable it is that such a reaction happens within the time interval Δt . Apparently, the probability is proportional to the time interval, thus it will be of the form $p_{\text{inf}}^{(S,I)} \Delta t$. Furthermore, the more infecteds are in the system, the more chances there are for one infected to pass on the disease—in fact, if every infected person has an equal chance of meeting other people, the probability of an infection occurring will be proportional to the number of infecteds,

$$p_{\text{inf}}^{(S,I)} \propto I.$$

For an infection to occur, it is not only necessary that an infected person meets somebody, but the person must meet a susceptible person. Recalling the assump-

tion of the well-mixed population, chances of meeting any person are equally distributed and thus the chance of meeting a susceptible person equals the proportion of this class in relation to the population size:

$$p_{\text{inf}}^{(S,I)} \propto I \frac{S}{N}.$$

Finally, we want to write an equal sign and give a name to the (unknown) proportionality constant:

$$p_{\text{inf}}^{(S,I)} = \alpha \frac{S I}{N}.$$

This proportionality constant is the reaction rate. Why? Given one infected person in a large¹ and fully susceptible population ($p_{\text{inf}}^{(S,I)} \Delta t = \alpha \Delta t$), α states the probability of an infection within a unit time interval. But we should be aware of what we have done: We stopped modeling the infection process, i.e., we stopped thinking about what happens in reality beyond meeting other people and put everything else into this single proportionality constant.

Similarly, but much more straightforward, we can formulate the probability of S being increased, which can only happen by a recovery reaction ($I \rightarrow S$). Since this is a spontaneous reaction which happens at rate β , the chances for a single infected person to recover within the time interval Δt are $\beta \Delta t$. Since there are I infected persons in the population, the probability that any one infected person of the population recovers, i.e., the probability of changing to $(S + 1, I - 1)$ from (S, I) within the time interval Δt , is

$$p_{\text{rec}}^{(S,I)} \Delta t = \beta I \Delta t.$$

Almost done. We need to assemble the parts: Given the initial condition (S_0, I_0) , the probability of finding S_0 susceptibles and I_0 infecteds in the population at time $t = 0$ is, of course, 1. But given the probability $P_{(S,I)}^t$ of finding the system in any particular state (S, I) at time t , what is the probability $P_{(S,I)}^{t+\Delta t}$ of finding the system in the same state at time $t + \Delta t$? We can write

$$P_{(S,I)}^{t+\Delta t} = P_{(S,I)}^t + \Delta P_{(S,I)}^t$$

and need to figure out $\Delta P_{(S,I)}^t$: How does the probability of finding the system in state (S, I) change within a small time interval? This we already investigated

¹Such that $\frac{S}{N} = \frac{N-1}{N}$ is essentially 1.

above, so we just need to collect the possible state transitions and do some book-keeping: All state transitions that go from (S, I) to some other state will decrease $P_{(S, I)}^{t+\Delta t}$ and all transitions that go to (S, I) will increase it. The transition $(S, I) \rightarrow (S-1, I+1)$, i.e., an infection, occurs with probability $p_{\text{inf}}^{(S, I)} \Delta t$ if the system is in state (S, I) , which is fulfilled with probability $P_{(S, I)}^t$. Thus, the probability of the infection event to happen is $P_{(S, I)}^t p_{\text{inf}}^{(S, I)} \Delta t$. Similarly, the probability of the infection event $(S+1, I-1) \rightarrow (S, I)$ is $P_{(S+1, I-1)}^t p_{\text{inf}}^{(S+1, I-1)} \Delta t$. The latter infection event increases the probability $P_{(S, I)}^{t+\Delta t}$ while the former decreases it. Similar considerations can be made for the recovery events and we will end up with the following equation:

$$\begin{aligned} P_{(S, I)}^{t+\Delta t} &= P_{(S, I)}^t \\ &+ P_{(S+1, I-1)}^t p_{\text{inf}}^{(S+1, I-1)} \Delta t + P_{(S-1, I+1)}^t p_{\text{rec}}^{(S-1, I+1)} \Delta t \\ &- P_{(S, I)}^t p_{\text{inf}}^{(S, I)} \Delta t - P_{(S, I)}^t p_{\text{rec}}^{(S, I)} \Delta t. \end{aligned}$$

We can take the limes $\Delta t \rightarrow 0$ and obtain an ordinary differential equation (throughout this thesis I will use d_t as a shortcut for $\frac{d}{dt}$):

$$d_t P_{(S, I)}^t = P_{(S+1, I-1)}^t p_{\text{inf}}^{(S+1, I-1)} + P_{(S-1, I+1)}^t p_{\text{rec}}^{(S-1, I+1)} - P_{(S, I)}^t (p_{\text{inf}}^{(S, I)} + p_{\text{rec}}^{(S, I)}). \quad (2.1)$$

This is the so-called *master equation* of the SIS model.² It governs the complete time evolution of the probability of finding the system in state (S, I) , for any state (S, I) . With a given initial condition, the dynamics of the system are determined.

²In its general form, the master equation governs the dynamics of stochastic systems. Such systems are represented by K possible states and a $K \times K$ transition probability rate matrix T which entries t_{kl} state the probability rate for the transition from state l to state k . The dynamical variables are the probabilities P_k ($k = 1, \dots, K$) of finding the system in state k and the evolution thereof is described by

$$d_t P_k = \sum_l P_l t_{kl} - P_k \sum_l t_{lk}.$$

There is a continuous version as well: If the system cannot be modeled by discrete states, functions $p(y|x)$ must be given for the probability rate at which the system transits from state x to state y and the evolution of the probability of finding the system in state x is governed by

$$d_t P(x) = \int P(y) p(y|x) dy - P(x) \int p(x|y) dy.$$

More on master equations and other random stuff can be found in Gardiner's *Handbook of Stochastic Methods* [Gar85].

Well, actually, the stochasticity of the dynamics is determined—the actual evolution of an actual system is then of course subject to random fluctuations.

This might appear somewhat unsatisfactory since it is hard to actually tell something about the dynamics of the SIS model. There is one thing, however: If there are no infecteds in the system, the infection probability rate $p_{\text{inf}}^{(S, I)}$ vanishes ($p_{\text{inf}}^{(S, I)} \propto I$). Thus, if at any time the number of infecteds in the system drops to zero, it will never increase again. And, interestingly, since $I > 0$ implies a non-zero recovery probability rate $p_{\text{rec}}^{(S, I)}$, there will always be the possibility of the system going into this so-called *absorbing state* $(S, I) = (N, 0)$. Everything that is possible will also happen if one waits long enough, so we can conclude this chapter with the interesting insight that, in theory, every SIS-like disease in a single well-mixed population will eventually die out (the same applies to the SIR model—though for different reasons, as the interested reader will quickly figure out).

No, wait. By simulation we can gain a little more insight. Figure 2.1 on the following page visualizes the time evolution of the probability $P_{(S, I)}$ for different model parameters. It can easily be seen that for small α , i.e., less “aggressive” diseases, and small system sizes the probability of the absorbing state quickly increases. Thus, in these cases the disease is expected to go extinct very fast. For large α , or to be more precise, for any α but large enough systems there seems to exist a stationary solution. From the above considerations we know that $\lim_{t \rightarrow \infty} P_{(S, I)}^t = \delta_{I, 0}$ since every trajectory will eventually hit the absorbing state, but the time scale of the decay increases with the system size and, in fact, for half of the simulated systems the decay was slower than the numerical accuracy of the simulation program was able to capture. Also note that with increasing system size the appearingly stationary state approaches the value marked by the dotted lines. And last but not least, notice that the relative variance of the $I > 0$ branch of the probability distribution shrinks with increasing system size. In the limit $N \rightarrow \infty$ we would expect the relative variance, i.e., the influence of the stochastic fluctuations, to vanish.

2.2 Ignoring Chances: The Mean-field Approximation

Whenever one encounters statistics it is worth taking a look at the statistical moments. The same is true for the dynamics of the stochastic compartmental models introduced above—which are described by time-evolving probability distributions.

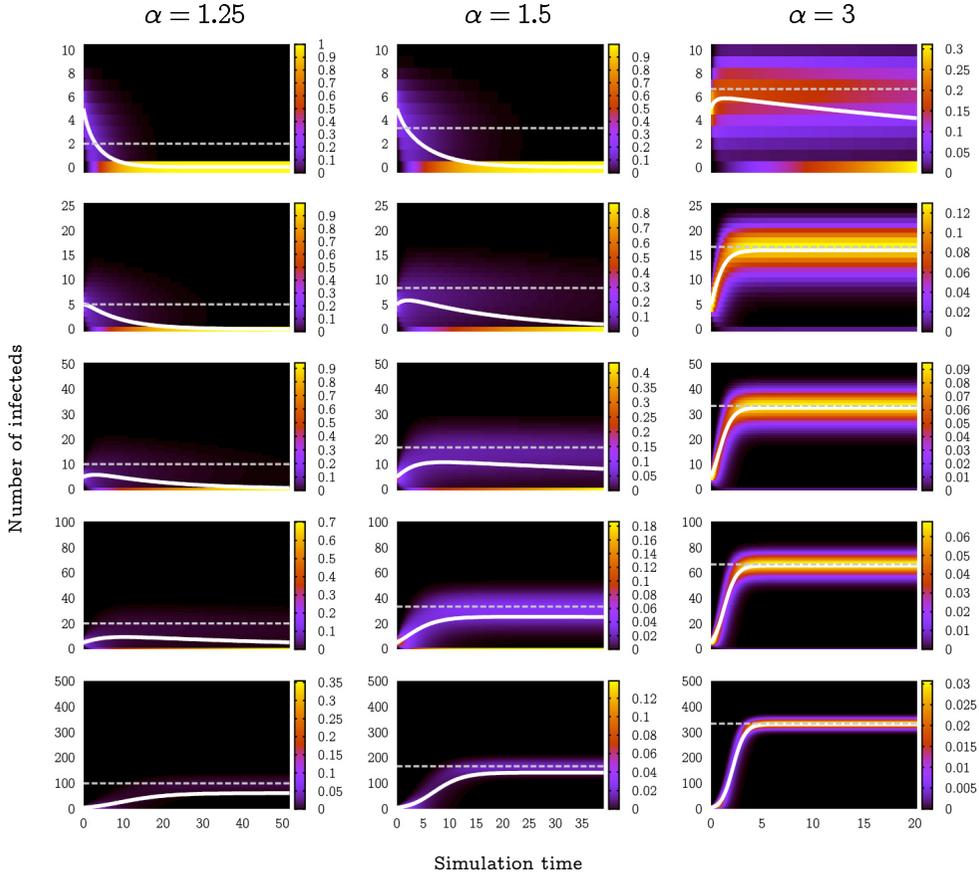


Figure 2.1: Time evolution of the SIS model's state probability distribution. Simulation time runs along the x -axis and the state index (number of infecteds) along the y -axis. The colors code the probability of finding the system in state (S, I) at time t , $P_{(S,I)}^t$. The simulations have been carried out for $\beta = 1$, three different values of α , and five different system sizes $N = 10, 25, 50, 100, 500$ (from top to bottom). The initial condition was $P_{(S,I)}^{t=0} = \delta_{I,5}$, i.e., the system was set up with five infected individuals. The color band range was chosen to be bright yellow for the maximum value of $P_{(S,I)}^t$ at the largest simulated time, i.e., $\max_I P_{(S,I)}^{t_{\max}}$. The solid white lines are the mean values $\langle I \rangle(t)$, the dotted grey lines mark the value $I^* = N(1 - \beta/\alpha)$.

The most interesting quantity we can look at is the first moment $\langle I \rangle$, the *expected* number of infecteds in the system at a time t . Here is the differential equation that governs its evolution:

$$\begin{aligned} d_t \langle I \rangle = d_t \sum_{I=0}^N P_{(S,I)}^t I = \sum_{I=0}^N \left(P_{(S+1,I-1)}^t p_{\text{inf}}^{(S+1,I-1)} + P_{(S-1,I+1)}^t p_{\text{rec}}^{(S-1,I+1)} \right. \\ \left. - P_{(S,I)}^t \left(p_{\text{inf}}^{(S,I)} + p_{\text{rec}}^{(S,I)} \right) \right) I. \end{aligned}$$

Nothing exciting happened so far—we simply inserted Eq. (2.1). Note, however, that the sum over all states is parameterized by I only. This is possible due to the conservation law $N = S + I$: For any given I the number of susceptibles is defined and thus the complete state of the system (S, I) can be described by one variable—either I or S . This would not be possible in the SIR model, since the conservation law in that model reads $N = S + I + R$ and one needs two variables to uniquely identify the system's state. Let us exploit the conservation law again by expressing every S in the formula in terms of N and I :

$$\begin{aligned} d_t \langle I \rangle = \sum_{I=1}^N P_{(N-I+1,I-1)}^t p_{\text{inf}}^{(N-I+1,I-1)} I + \sum_{I=0}^{N-1} P_{(N-I-1,I+1)}^t p_{\text{rec}}^{(N-I-1,I+1)} I \\ - \sum_{I=0}^N P_{(N-I,I)}^t \left(p_{\text{inf}}^{(N-I,I)} + p_{\text{rec}}^{(N-I,I)} \right) I. \end{aligned}$$

Note also that we dropped two terms, the one for $I = 0$ in the first sum and the one for $I = N$ in the second sum. Both terms vanish because they involve the probabilities $P_{(N+1,-1)}^t$ and $P_{(-1,N+1)}^t$, which are zero for all times since they refer to unreachable (i.e., non-existing) system states. Next, we will express everything in terms of $P_{(S,I)}^t$ by rearranging sum indices:

$$\begin{aligned} d_t \langle I \rangle = \sum_{I=0}^{N-1} P_{(N-I,I)}^t p_{\text{inf}}^{(N-I,I)} (I+1) + \sum_{I=1}^N P_{(N-I,I)}^t p_{\text{rec}}^{(N-I,I)} (I-1) \\ - \sum_{I=0}^N P_{(N-I,I)}^t \left(p_{\text{inf}}^{(N-I,I)} + p_{\text{rec}}^{(N-I,I)} \right) I. \end{aligned}$$

We can safely add the $I = N$ term to the first sum because the infection probability rate is zero for $I = N$ (since then $S = 0$). Similarly, the $I = 0$ term can be added to the second sum, since the recovery probability rate vanishes for the state where

no infecteds exist. Then, the third sum cancels the $(I + 1)$ and $(I - 1)$ factors in the first and second sum, yielding the following:

$$d_t \langle I \rangle = \sum_{I=0}^N P_{(N-I,I)}^t p_{\text{inf}}^{(N-I,I)} - \sum_{I=0}^N P_{(N-I,I)}^t p_{\text{rec}}^{(N-I,I)}.$$

This is refreshingly simple and can be rewritten using the notation of expectation values as

$$d_t \langle I \rangle = \langle p_{\text{inf}}^{(S,I)} \rangle - \langle p_{\text{rec}}^{(S,I)} \rangle,$$

or, after inserting the definitions of the probability rates,

$$d_t \langle I \rangle = \alpha \frac{\langle SI \rangle}{N} - \beta \langle I \rangle.$$

Yet we still have not expressed everything solely in terms of $\langle S \rangle$ and $\langle I \rangle$ and still depend on other statistical properties than just the first moment. We have to resolve the correlate $\langle SI \rangle$ into simple expectation values and things would be simple if S and I were independent variables—but, of course, they are not, since $S = N - I$ (or $S = N - I - R$ in the SIR and SIRS model). However, we can make some approximation here. We would like to see $\langle SI \rangle \approx \langle S \rangle \langle I \rangle$, i.e.,

$$\langle (N - I) I \rangle = N \langle I \rangle - \langle I^2 \rangle \approx N \langle I \rangle - \langle I \rangle^2 = \langle N - I \rangle \langle I \rangle,$$

which boils down to approximate $\langle I^2 \rangle \approx \langle I \rangle^2$. This is valid if the variance of the probability distribution of finding a certain number of infected people is zero, that is, if there are no stochastic fluctuations in the number of infecteds.³

Of course, that is not the case, but the fluctuations will become less important when N is large. Figure 2.1 can be seen as empirical evidence for this claim, but I will try to present a more rigorous argument as well: The SIS system reflects a Markovian jump process on the I -axis where a jump to the left (smaller I) occurs with probability rate βI and a jump to the right (larger I) with rate $\alpha S I / N = \alpha (N - I) I / N$ (cf. Fig. 2.2). There is a state I^* where the two rates are equal:

$$\beta I^* = \alpha \frac{(N - I^*) I^*}{N} \Leftrightarrow I^* = N \left(1 - \frac{\beta}{\alpha} \right).$$

For $I > I^*$ the recovery rate βI is larger than the infection rate and vice versa for $I < I^*$. Thus, for $I \neq I^*$, the jump process is biased towards the state I^* . Assume

³The variance $\sigma_I^2 = \langle (I - \langle I \rangle)^2 \rangle = \langle I^2 - 2I \langle I \rangle + \langle I \rangle^2 \rangle = \langle I^2 \rangle - 2 \langle I \rangle^2 + \langle I \rangle^2$ is zero if and only if $\langle I^2 \rangle = \langle I \rangle^2$.

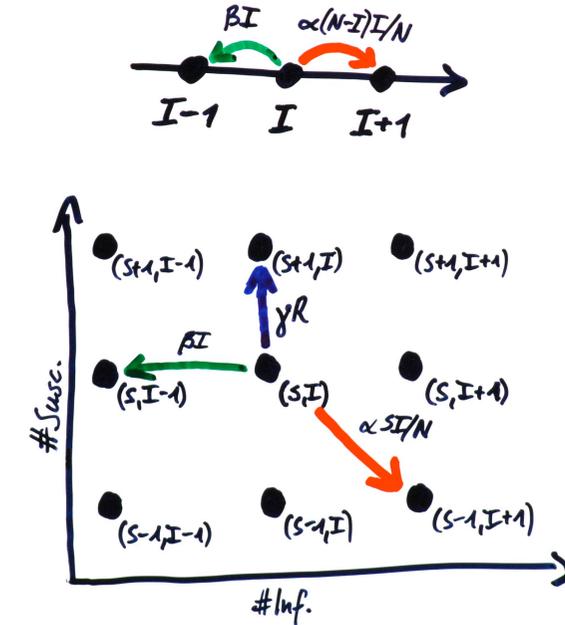


Figure 2.2: Jump processes in the state space. The top figure shows the state space of the SI and SIS model. Since the two dynamical variables S and I are coupled by the conservation law $N = S + I$, the state space is one-dimensional. An infection (red) leads to an increase of the number of infected people, I , while a recovery (green, only in SIS model) leads to a decrease. In the bottom figure, the state space of the SIR and SIRS is shown. Since here we have three dynamical variables and the conservation law $N = S + I + R$, two variables are necessary to index a state. The deimmunization event (blue) depicting the reaction $R \rightarrow S$ is present in the SIRS model only.

for a moment that for $I \approx I^*$ the probability rates are essentially equal to βI^* and $\alpha(1 - I^*/N)I^*$. Then, in the vicinity of I^* the jump process is an unbiased continuous time random walk with jump rate $w \approx \beta I^*$ in both directions. As I will derive more detailed in Sec. 2.5, if we start a random walk at I^* , after time t the system's *expected displacement* is \sqrt{wt} states. Now comes the interesting part: If we increase N , then w will increase proportionally. Why? Because $I^* \propto N$ and w was defined to be the jump rate in the vicinity of I^* . Thus, the expected displacement of the system after time t scales with \sqrt{N} . Now, the mindful reader will object that the assumption of equal jump rates around I^* is a bad one. In fact, the bias of the jump process in the vicinity of I^* , $|\beta I - \alpha(1 - I/N)I|$, also scales with N . Therefore, we can cheerfully conclude that \sqrt{wt} is an *upper bound* for the expected displacement of the system (due to the bias in the jump rates, one would probably expect a time-independent upper bound as well, but this result is enough for our purpose). So, the expected displacement, i.e., the fluctuations of the system, scale with \sqrt{N} . This means that for $N \rightarrow \infty$ the fluctuations increase unbounded as well, but their *relative* value \sqrt{N}/N , which determines their importance, goes to zero.

Why is the relative value interesting? The fluctuations influence the expected number of infecteds, $\langle I \rangle$, only if they are asymmetric around $\langle I \rangle$. This is the case due to the absorbing state $I = 0$ (and the finite system size that prohibits $I > N$). However, the fluctuations are driven away from this absorbing state when I^* scales linearly with N since the fluctuations scale sub-linearly. Thus, in the limit $N \rightarrow \infty$ we obtain

$$d_t \langle I \rangle = \alpha \frac{\langle S \rangle \langle I \rangle}{N} - \beta \langle I \rangle,$$

the so-called *mean-field approximation*⁴ of the SIS model, describing the time evolution of the means of the dynamical quantities. We can further unclutter notation and generalize the equation by substituting expected population numbers with the (expected) fraction of susceptibles $s := \langle S \rangle / N$ and infecteds $j := \langle I \rangle / N$:

$$d_t j = \alpha s j - \beta j \quad d_t s = d_t(1 - j) = -\alpha s j + \beta j \quad (2.2)$$

Similarly, the mean-field equations for the SIR and SIRS model can be derived,

⁴The name *mean-field* originates from the *mean-field theory* (or *self-consistent field theory*) in statistical physics, which is used to solve many-body systems with complex interactions by replacing them with a system of approximatively equivalent average, effective interactions: the mean interaction field.

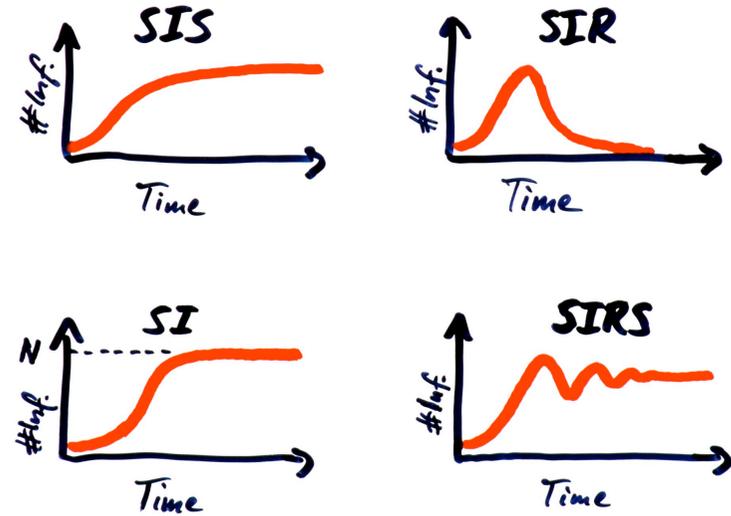


Figure 2.3: Qualitative mean-field behavior of the various compartmental models. The SIS and SIR model are the two models of interest in this study, but the SI and SIRS model are shown for reference as well. The SI model is very similar to the SIS except that, due to the lacking recovery reaction $I \rightarrow S$, the system will go into the fully infected state $(S, I) = (0, N)$ instead of a stable state (S^*, I^*) with $I^* < N$. The SIRS model extends the SIR model by an $R \rightarrow S$ reaction and exhibits damped oscillations around a fixed point.

which is left as an exercise to the interested reader:

$$\begin{aligned} d_t j &= \alpha s j - \beta j & d_t s &= -\alpha s j & d_t r &= \beta j & (\text{SIR}) \\ d_t j &= \alpha s j - \beta j & d_t s &= -\alpha s j + \gamma r & d_t r &= \beta j - \gamma r & (\text{SIRS}) \end{aligned}$$

Now, we can have a look at the dynamics of the expectation value. First, we easily determine the fixed points of the SIS model. Clearly, if $j = 0$, both $d_t j$ and $d_t s$ are zero. The other root of the function is determined by $\alpha s - \beta = 0$ and with the conservation law $s + j = 1$ we thus obtain the two fixed points

$$j_1^* = 0, \quad s_1^* = 1 \quad \text{and} \quad j_2^* = 1 - \frac{\beta}{\alpha}, \quad s_2^* = \frac{\beta}{\alpha}.$$

The first (trivial) fixed point is not surprising: It tells us that if the system lacks infected people it will stay like this. We already know this from our considerations

in the last section. The second (non-trivial) fixed point, however, is contrary to the results we found above. It tells us that there exists an equilibrium concentration of infected people, which is also called the *endemic state*—an SIS-like disease in a single population will maintain a constant concentration j_2^* of infecteds (remember: we concluded the previous section with the statement that every SIS-like disease will eventually go extinct). But it is important to note that the mean-field just describes the time evolution of the expectation value in the absence of fluctuations while the master equation describes the time evolution of the whole probability distribution. The mean-field description does not capture the effects of the absorbing state $I = 0$ and thus it is not surprising that it predicts a stable non-zero level of infected people, as the stochastic system is also biased towards $I^* = N j_2^*$.

For the SIS model, we can even solve the differential equations and obtain an analytical expression for the course of an epidemic:⁵

$$j(t) = \frac{1}{\left(\frac{1}{j_0} + \frac{\alpha}{\alpha - \beta}\right) \exp(-(\alpha - \beta)t) + \frac{\alpha}{\alpha - \beta}}.$$

For $t \rightarrow \infty$ the exponential vanishes and $j(t) \rightarrow (\alpha - \beta)/\alpha = j_2^*$ as we already know from the fixed point analysis. Note that for $j_0 = 0$ the solution is undefined—in this case $j(t) \equiv 0$ due to the trivial fixed point $j_1^* = 0$.

The SIR model cannot be solved analytically, but we can learn about the dynamics from a fixed point analysis. Clearly, there is a fixed point

$$j^* = 0,$$

⁵In equation (2.2), we substitute $\nu = 1/j$ and $d_t \nu = d_t(1/j) = -d_t j / j^2$ (thus $d_t j = -d_t \nu / \nu^2$) and obtain

$$-\frac{1}{\nu^2} d_t \nu = \alpha \left(1 - \frac{1}{\nu}\right) \frac{1}{\nu} - \beta \frac{1}{\nu},$$

hence $d_t \nu = -(\alpha - \beta)\nu + \alpha$.

This we can solve by the ansatz

$$\nu(t) = A \exp(-Bt) + C, \quad d_t \nu(t) = -AB \exp(-Bt).$$

By insertion we get

$$-AB \exp(-Bt) = -(\alpha - \beta)(A \exp(-Bt) + C) + \alpha$$

and see immediately that $C = \alpha/(\alpha - \beta)$, $B = \alpha - \beta$ and that A is to be determined by the initial condition $\nu(t=0) = \nu_0$:

$$\nu(t=0) = A + C = \nu_0 \quad \Leftrightarrow \quad A = \nu_0 + \frac{\alpha}{\alpha - \beta}$$

Substituting back (with $j_0 = \frac{1}{\nu_0}$) we obtain the solution for Eq. (2.2).

regardless of the values s^* and r^* . Also, for $s = \beta/\alpha$ and $j > 0$ we find $d_t j = 0$ as in the SIS model, but it is not a fixed point since $d_t s < 0$ and $d_t r > 0$. This is interesting, since $d_t j$ changes its sign depending on s . Thus, while the trivial fixed point is unstable for $s > \beta/\alpha$, i.e., $d_t j > 0$ for $j > 0$, it will become stable if $s < \beta/\alpha$, since then the number of infecteds is driven back to zero.⁶ Thus, if we start with a fully susceptible population and a small amount of infecteds $j_0 > 0$, the density of infecteds will rise until the density of susceptibles has dropped to $s = \beta/\alpha$ (note that at that time $j \neq 1 - \beta/\alpha$, since some of the infecteds will have recovered already). After that, the susceptibles will further decrease but the infection reactions do not outrun the recovery events anymore and thus, while the density of recovered further increases, the density of infecteds decrease. For $t \rightarrow \infty$, the disease will have died out and the density of recovered will have approached a fixed point r^* with $s^* = \rho - r^*$, which values cannot be computed analytically. The qualitative dynamics of the SIR model are depicted in Fig. 2.3 (top right).

2.3 Megacities and Ghost Towns: Variable Population Density

In this section, we will shortly revisit the compartmental models introduced in Sec. 2.1 and extend the concept to situations where population density plays a role. Remember that we deduced the infection probability rate

$$p_{\text{inf}}^{(S,I)} = \alpha \frac{SI}{N}$$

from the two considerations that it should be proportional to the number of infecteds, I , and the probability that an infected person meets a susceptible person, S/N . We have implicitly assumed that the population density is 1. If the population is spread over an area Ω , one would expect the probability of one person meeting another person to scale with the inverse of the area, if the population size

⁶The observing reader may object that even for large s , the number of infecteds will only raise if $\alpha > \beta$. This is right and an important insight that I completely overlooked above. We can define the *basic reproduction number* $R_0 = \alpha/\beta$ as the average number of people an infected person in a fully susceptible population will infect during its infection period (remember that α can be interpreted as the average number of infectious contacts and β as the recovery rate, thus β^{-1} as the average time of infection). Only if $R_0 > 1$ a disease breakout is possible, since otherwise not enough secondary cases are produced before everybody is healthy again.

is constant, and linearly with the population size, if the area is constant, thus:

$$p_{\text{inf}}^{(S,I)} = \alpha I \frac{S}{N} \frac{N}{\Omega} = \alpha \frac{S I}{\Omega}.$$

This naturally leads to the mean-field equation for the *density of infecteds* $j := I/\Omega$ (with density of susceptibles $s := S/\Omega$ and total population density $\rho := N/\Omega$):

$$d_t j = \alpha s j - \beta j.$$

However, if we want to stick to the notion of *fraction of infecteds* $j = I/N$, the equation reads

$$d_t j = \alpha \rho s j - \beta j.$$

The former equation looks similar to the equation we already derived in Sec. 2.2 but the fixed point is different because now $s + j = \rho$ and thus

$$j_1^* = 0, \quad j_2^* = \rho - \frac{\beta}{\alpha},$$

i.e., the fixed point scales linearly with the population density. This might not be surprising because we would suspect to have higher contact rates in dense populations and therefore a higher endemic state. Nonetheless, it is remarkable that the corresponding density of susceptibles is independent of the total population density, $s_2^* = \beta/\alpha$, while, of course, the equilibrium *fraction* of susceptibles does scale with population density: $s_2^* = \beta/(\alpha\rho)$. Also, it is worth noting that the fixed point ceases to exist if $\rho < \beta/\alpha$ (since there cannot be a negative density of infected people) and thus there is a so-called *epidemic threshold* on the population density below which the disease cannot break out.

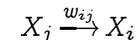
2.4 Population Networking: Spatially Extended Systems

So far, we have seen models that can predict the course of a disease brought into a single well-mixed population. Obviously, these models' applications are very restricted. The world is not a single well-mixed population but individuals interact in a very complex way with each other. The next step in the modeling process towards a more versatile model is to set up a system of multiple populations. Each of the populations is assumed to be well-mixed, i.e., we simplify the complex interactions between the people within a single population by assuming that they have

equal contact probabilities. However, we introduce some sort of coupling between the populations that is not necessarily in the way that each population has equal contact probability with every other population. Thus, we have broken down the whole population of, e.g., a country, in which the assumption of equal contact probabilities is plainly wrong, into several populations of, e.g., cities, that are connected by travellers between the cities and in which the equal contact probability assumption is probably a bit less wrong.

The mathematical representation of this is a digraph $G = (P, C)$ which nodes P represent populations and which directed edges C represent couplings between populations (Fig. 2.4). The weights of the edges indicate the strength of the coupling.⁷ In the following, we will rethink everything we have done in the last sections and apply it to spatially extended systems.

Assume we have M populations with population sizes N_i ($i = 1, \dots, M$). Or, put in different words, we have $N = \sum_{i=1}^M N_i$ individuals that can be classified into M classes by their population membership. Travelling from one population to another can be expressed as the chemical reaction



where a person from population j becomes a person from population i .⁸ This reaction occurs at rate w_{ij} , which is the coupling strength mentioned above. From this, we can construct a master equation for the probability P_i^t of a person finding

⁷There are two fundamentally different ways to model the interaction between two populations: With *dispersive coupling*, the coupling strength denotes some sort of travelling rate and individuals from one population diffuse to the other population (and then to another population and so on) according to that travelling rate. This would perfectly model the travel behavior of hobos, people who—especially in the time of the Great Depression—were moving from place to place looking for temporary work. If dispersive coupling is modified such that each individual has a home population where it returns before diffusing to a different population, it can be used to model the travel behavior of commuters. However, the actual travel behavior of actual people is probably somewhere in between. In contrast to the dispersive coupling, in *directly coupled systems* the coupling strength denotes the fraction of the population that is constantly interacting with the other population, i.e., it denotes the overlap between two populations in terms of the equal contact probability assumption. This is not modeling an actual travel behavior but rather an effective contact between populations. In this study, I focus on dispersive coupling since this is the only coupling for which we can derive a spatially continuous reaction-diffusion equation.

⁸Yes, I use j for a time-dependent function and for an integer index—later I will also use it to denote a time- and space-dependent function. It is a long tradition in physics to use ambiguous notation and the reader is expected to work out the meaning of each appearance from the context.

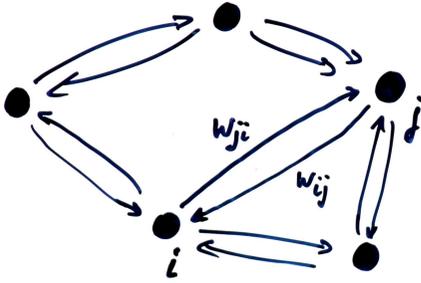


Figure 2.4: Network of populations with inter-population couplings. The nodes represent populations (or cities, countries etc.) and the weighted and directed links between them indicate an exchange between the populations (i.e., travelling).

itself in population i :

$$d_t P_i^t = \sum_{j=1}^M P_j^t w_{ij} - \sum_{j=1}^M P_i^t w_{ji}.$$

The expected number of people in population i is then given by $N_i = N P_i^t$ and its evolution by the mean-field equation

$$d_t N_i = \sum_{j=1}^M w_{ij} N_j - \sum_{j=1}^M w_{ji} N_i.$$

The first sum accounts for all people travelling from some other population into population i and the second sum accounts for all people travelling away from population i .

2.5 Indiscrete Models: From Compartments to Continuous Variables

The next step in the modeling process involves a transition to continuous variables in the spatial domain. Why would we want to do this? Back in the Middle Ages, the country was populated by many little villages which were home to small populations. Furthermore, long distance travelling was negligible and contacts

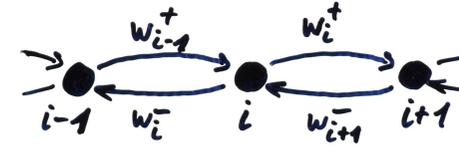


Figure 2.5: Linear chain of populations.

were essentially restricted to the neighboring villages. This can be modeled by setting up a network with a large number of nodes (populations), each of which holds a small number of people and is connected to its nearest neighbors. In the limit of infinite number of infinitesimally small populations in a finite area, the total population is not constrained to discrete nodes but spreads continuously over a one- or two-dimensional area. Travelling in this case will not be modeled by population flows from one node to another, but by diffusion.

The observing reader may object that the assumptions made in the beginning are not adequate anymore—and she would be perfectly right. Today’s populations are mostly concentrated into cities and thus the approach of geographically embedded network models seems to be much more appropriate. Furthermore, travelling behavior changed significantly as well. Contacts are not restricted to the neighboring populations anymore, but novel transportation means such as highways, intercity trains and especially intercontinental air traffic routes provide fast long-range travelling opportunities resulting in significant contact rates between distant populations which cannot be described by normal diffusion. Again, the network approach seems to be appropriate.

However, the diffusive continuous model can still serve as a sandbox—a “toy model” in which certain aspects of idealized scenarios can be investigated. In particular, it is easily possible to define the propagation speed of a pandemic wave crossing a country in a continuous model, which is difficult to do in an arbitrary network. That said, I already forestalled an important result: In continuous models, wave solutions exist and describe the time evolution of the spatial distribution of, e.g., the Black Death pandemics in the Middle Ages. But let us derive continuous models first and then take a look at their characteristics.

Consider a linear chain of populations with links between nearest neighbors (Fig. 2.5). In the network formulation, this is expressed by the coupling strengths

$$w_{ji} = w (\delta_{i,j-1} + \delta_{i,j+1}),$$

i.e., a link is established from population i to population j if j is either the preceding or succeeding population to j . In this formulation, all links are symmetric and of equal strength w . We might want to generalize this to

$$w_{ji} = w_i^- \delta_{i-1,j} + w_i^+ \delta_{i+1,j}.$$

Now the coupling of the i -th population to the preceding (“left”) population in the chain will be of strength w_i^- and the coupling strength to the $(i+1)$ -th (“right”) population will be w_i^+ . Following the mean-field consideration from the last section, the population dynamics will be governed by the ordinary differential equation

$$\begin{aligned} d_t N_i &= \sum_j w_{ij} N_j - \sum_j w_{ji} N_i \\ &= \sum_j (w_j^- \delta_{i,j-1} N_j + w_j^+ \delta_{i,j+1} N_j - w_i^- \delta_{j,i-1} N_i - w_i^+ \delta_{j,i+1} N_i) \quad (2.3) \\ &= w_{i+1}^- N_{i+1} + w_{i-1}^+ N_{i-1} - w_i^- N_i - w_i^+ N_i. \end{aligned}$$

We did not yet embed the chain into a geographical space. To do this, we simply associate with each node a position on a one-dimensional line such that node i is at position $x_i = i \Delta x$ where Δx denotes the distance between two adjacent populations. Furthermore, we have to sacrifice some of the generality regarding the coupling strength introduced above in order to be able to derive a diffusion equation: we will assume a symmetric coupling of the form $w_i^- = w_i^+ =: w_i$ (note that the symmetry is around the nodes—the network’s links can still be asymmetric). Now, to give the geographical information any relevant meaning, it has to be found in the dynamical variables—this is another step in the modeling process which has to be motivated phenomenologically. Technically, we will introduce a completely new model, though based on the same ideas, in which the old model reappears as the special case $\Delta x = 1$.⁹ First, we would assume that each node’s population is proportional to the space it is assigned: $N_i = n_i \Delta x$. Here, we implicitly introduced the spatial population density n_i . Furthermore, we would suspect that the coupling strength should be anti-proportional to the squared distance between two populations: $w_i = D_i / \Delta x^2$. While the first is obvious, the latter might require a more elaborated motivation.

⁹Which is, according to Albert Einstein, the most joyful thing that can happen to a theory: “Es ist das schönste Los einer physikalischen Theorie, wenn sie selbst zur Aufstellung einer umfassenden Theorie den Weg weist, in welcher sie als Grenzfall weiterlebt.” (*Über die spezielle und die allgemeine Relativitätstheorie (gemeinverständlich)*, written 1916, published 1972, Friedr. Vieweg & Sohn, Braunschweig, p. 47)

For this, let me further simplify the system to equal coupling strengths $w_i = w$. The results are valid for any couplings that are symmetric around the nodes, as the observing reader will intuitively notice, but the motivation is much easier to present when restricted to uniform couplings.

People are performing a continuous time random walk on the (one-dimensional) lattice of cities: they jump from one city to one of its neighbors in the (infinitesimal) time interval dt with probability $2w dt$ (there are two links of strength w). It can easily be shown that the expected displacement of a single person after k random walk steps is proportional to \sqrt{k} lattice sites.¹⁰ This corresponds to a spatial distance $X_k = \Delta x \sqrt{k}$ since the lattice sites are Δx apart from each other. But we are not interested in the expected displacement after k steps but rather in the expected displacement after some time t . So, how does k relate to t ? We can easily calculate the time t_k needed for k steps from the so-called *waiting times* $\tau_k = t_k - t_{k-1}$ between two jumps:

$$t_k = \sum_{i=1}^k \tau_k.$$

The waiting times are random and, therefore, so is t_k , but if we knew the expectation value of τ_k we could calculate the expected time $\langle t_k \rangle$ needed for k jumps. It can be shown¹¹ that the waiting times are exponentially distributed with $\langle \tau \rangle = (2w)^{-1}$ and thus

$$\langle t_k \rangle = \sum_{i=1}^k \langle \tau_k \rangle = k (2w)^{-1}.$$

¹⁰Consider a (discrete time) random walk process on a lattice where the walker performs a jump to the left or to the right with equal probability in each time step. Starting at $x = 0$, the position of the walker after k steps is $x_k = \sum_{i=1}^k \xi_k$ where ξ_k is the distance moved in jump k , either 1 or -1 , both with probability $\frac{1}{2}$. The expected position after k steps is $\langle x_k \rangle = \sum_{i=1}^k (\frac{1}{2}(1) + \frac{1}{2}(-1)) = 0$, which is not surprising. More interesting is the second moment of the position, $\langle x_k^2 \rangle = \sqrt{\sum_{i=1}^k \langle \xi_k^2 \rangle} = \sqrt{\sum_{i=1}^k 1} = \sqrt{k}$. It indicates the “range” of the random walker: the larger the second moment, the further away from the starting position you can find the walker without being substantially surprised. It is therefore also referenced to as the *expected displacement* after k steps.

¹¹Let $P(\tau) dt$ be the probability that a jump occurs in the (infinitesimal) time interval $[t + \tau, t + \tau + dt)$. Since $2w dt$ is the probability for a jump occurring in a time interval of length dt , we can write $P(\tau) dt = P_0(\tau) 2w dt$ where $P_0(\tau)$ is the probability that a jump does *not* occur within $[t, t + \tau)$. Dividing this interval into K small pieces of length ε , in each of which the probability of *not* jumping is $1 - 2w\varepsilon = 1 - 2w\tau/K$, we obtain an expression for $P_0(\tau)$ and $P(\tau)$ by calculating the joint probability of not jumping in *any* of the K pieces and taking

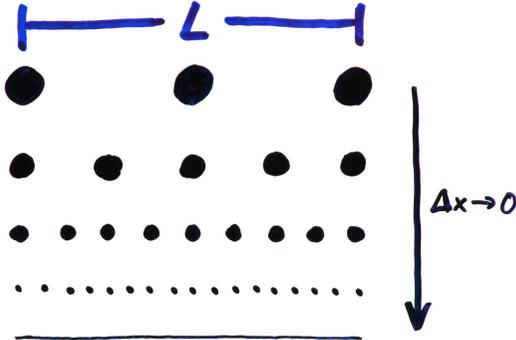


Figure 2.6: The limit of fitting an infinite number of infinitesimally small populations into a finite space.

Remember the expected displacement after k steps, $X_k = \Delta x \sqrt{k}$, which we can now write as

$$X(t) = \Delta x \sqrt{2 w t}.$$

This expected displacement should be invariant under changes in Δx : Redistributing people within the length L by changing the underlying lattice should not hinder any person to travel the distance $X(t)$ within time t —although it would not be the same number of jumps. Thus, for two different lattice spacings Δx_1 and Δx_2 and two travel rates w_1 and w_2 the relation $\Delta x_1 \sqrt{w_1} = \Delta x_2 \sqrt{w_2}$ must hold and thus

$$w_2 = w_1 \left(\frac{\Delta x_1}{\Delta x_2} \right)^2.$$

If we start with a linear chain of cities each $\Delta x_1 = 1$ apart and then change to a linear chain of cities with distance $\Delta x_2 = \Delta x$, we have to replace the travel rates

the limit $K \rightarrow \infty$:

$$P_0(\tau) = \lim_{K \rightarrow \infty} \left(1 - \frac{2 w \tau}{K} \right)^K = \exp(-2 w \tau),$$

$$P(\tau) = 2 w \exp(-2 w \tau).$$

For a more general derivation of this result see [Gil76].

w_i with $w_i/(\Delta x)^2$:

$$d_t n_i = \frac{1}{(\Delta x)^2} (w_{i+1} n_{i+1} - 2 w_i n_i + w_{i-1} n_{i-1}).$$

In the limit $\Delta x \rightarrow 0$ (Fig. 2.6) we obtain continuous variables $n(x)$ and $w(x)$ and the right hand side will converge to the second spatial derivative—resulting in a diffusion equation. For conventional reasons, and to make it easier to distinguish between the diffusion coefficient in the continuous system and the travel rates in the discrete, I define $D(x) := w(x)$:

$$d_t n = \partial_{xx}(D n).$$

Up to now we have restricted ourselves to couplings that are symmetric around the nodes. We will now relax this constraint by reconsidering biased couplings: w_i^+ , the coupling from city i to $i + 1$, and w_i^- , the coupling from city i to $i - 1$. We can introduce two new variables,

$$D_i := \min \{w_i^+, w_i^-\}, \quad F_i := w_i^+ - w_i^-.$$

The idea behind those definitions is to separate the travel events into undirected diffusion and directed drift. The first, D_i , is the travel rate out of city i disregarding direction, or, in other words, it is the number of travel events per time unit in one direction, for which there is a compensating travel event into the other direction. D_i is thus the “diffusive part” of the travel behavior. The second, F_i , is a measure for the “directedness” of the coupling around city i . More specific, $|F_i|$ is the probability rate for travel events that go to the right (if $F_i > 0$) or to the left (if $F_i < 0$) for which no compensating travel event in the opposite direction occurs.

We can reconstruct the original coupling strengths from our new variables, but we have to distinguish two cases:

$$w_i^+ = \begin{cases} D_i + F_i & \text{if } F_i > 0 \\ D_i & \text{otherwise} \end{cases}$$

$$w_i^- = \begin{cases} D_i & \text{if } F_i > 0 \\ D_i - F_i & \text{otherwise} \end{cases}$$

Inserted into the discrete mean-field equation (2.3), we obtain for the case $w_i^+ > w_i^-$ ($F_i > 0$):

$$d_t N_i = (D_{i-1} + F_{i-1}) N_{i-1} - D_i N_i - (D_i + F_i) N_i + D_{i+1} N_{i+1}$$

$$= (D_{i-1} N_{i-1} - 2 D_i N_i + D_{i+1} N_{i+1}) - (F_i N_i - F_{i-1} N_{i-1}).$$

In case $w_i^+ < w_i^-$ we have

$$d_t N_i = (D_{i-1} N_{i-1} - 2 D_i N_i + D_{i+1} N_{i+1}) - (F_{i+1} N_{i+1} - F_i N_i).$$

How does this equation look like in the continuous limit? Similarly to the approach taken above, we can substitute population numbers N_i with densities n_i . Since the D_i are to be interpreted as effective diffusion travel rates, we can replace D_i with $D_i/(\Delta x)^2$ as motivated above. Also, in a similar reasoning we can motivate that F_i should be replaced with $F_i/\Delta x$. Why does Δx come in linearly instead of squared? Assume a homogeneous linear chain with $w_i^- = 0$ and $w_i^+ = w$. There is no normal diffusion in this system and all movements are governed by the directed $F_i = w$. Thus, after k steps, a person has traveled the distance $X_k = k \Delta x$. Similarly, after time t , the person has traveled the distance $X(t) = \Delta x w t$. Obviously, when decreasing Δx , w has to be increased proportionally in order to conserve $X(t)$. Thus, the mean-field equation for arbitrary lattice spacings reads

$$d_t n_i = \frac{1}{(\Delta x)^2} (D_{i-1} n_{i-1} - 2 D_i n_i + D_{i+1} n_{i+1}) - \frac{1}{\Delta x} (F_{i+1} n_{i+1} - F_i n_i),$$

which in the limit $\Delta x \rightarrow 0$ becomes the general *Fokker-Planck equation* consisting of a diffusion term already seen above and a *drift term* that accounts for the asymmetry of the coupling strengths,

$$\partial_t n = \partial_{xx}(Dn) - \partial_x(Fn). \quad (2.4)$$

So, now we have a nice and shiny equation describing the time evolution of the distribution of people spread over an infinite one-dimensional space for given local diffusion coefficients $D(x)$ and local drift coefficients $F(x)$. But we have somehow degraded to a post-revolutionary kind of society where everybody is equal—though this might be compelling for politicians and human rights activists, this is certainly counterproductive from the epidemiological point of view. The solution to combine the society of classes with a spatially continuous metapopulation model is intuitively obvious—with the local densities of susceptibles, $s(x)$, and infecteds, $j(x)$, the *spatially continuous one-dimensional mean-field SIS model* is given by these two partial differential equations:

$$\begin{aligned} \partial_t s &= -\alpha s j + \beta j + \partial_{xx}(D s) - \partial_x(F s), \\ \partial_t j &= +\alpha s j - \beta j + \partial_{xx}(D j) - \partial_x(F j). \end{aligned} \quad (2.5)$$

In general, the dynamics of a spatially continuous compartmental mean-field model is described by

$$\partial_t \vec{u} = R(\vec{u}) + \partial_{xx}(\vec{D} \vec{u}) - \partial_x(\vec{F} \vec{u}),$$

where $\vec{u}(x)$ is the vector of local densities of different population classes, $R(\vec{u})$ is the *local reaction term*, also referenced to as the *local interaction term*, and the vector symbol above the diffusion and force coefficients indicate that it is possible to have different diffusion and force strengths affecting the different types of individuals (we will indeed exploit this differentiation later).

2.5.1 Travelling Waves in Spatially Continuous Systems

Now that we have found some nice and compact analytical models, let us dissect them. In the following, we will assume that in the SIS model with global diffusion and no drift there exists a stable wave solution, i.e., the function $j(x, t)$ describing the spatio-temporal evolution of the density of infecteds is in fact a function of one variable of the form $j(\xi)$ with $\xi = x - ct$ where c is the wave front velocity. The partial derivatives of this function can then be substituted,

$$\partial_t j \rightarrow -c d_\xi j, \quad \partial_{xx} j \rightarrow d_{\xi\xi} j,$$

and the second of the equations in (2.5) (with and $D(x, t) = D_0$ and $F(x, t) = 0$) becomes a second-order ordinary differential equation:¹²

$$-c d_\xi j = \alpha(\rho - j)j - \beta j + D_0 d_{\xi\xi} j.$$

We now define $\iota(\xi) := d_\xi j$ and obtain the following set of first-order ODEs of two variables:¹³

$$d_\xi j = \iota, \quad d_\xi \iota = -\frac{c}{D_0} \iota - \frac{\alpha(\rho - j)j - \beta j}{D_0}.$$

What do these two variables represent in physical terms? By construction, $j(\xi)$ is the *wave form* of the travelling wave, since for a fixed t , e.g. $t = 0$, it is the population density profile $j(x)$. Also, it describes the local time evolution of the density of infected people, since for a fixed x , e.g. $x = 0$, $j(\xi) = j(-ct)$. Note that $j(\xi)$ is describing the backwards evolution (the argument is $-t$), we therefore expect $\lim_{\xi \rightarrow \infty} j(\xi) = 0$ and $\lim_{\xi \rightarrow -\infty} j(\xi) = j^*$, i.e., a location will have no infecteds at $t \rightarrow -\infty$, will be infected by people diffusing into this location and end up with the equilibrium density j^* at $t \rightarrow \infty$.

¹²In the following, I use ρ again rather than n to denote the total population density.

¹³The Greek *iota* is pronounced with a j at the beginning, thus ι is the best choice to denote something related to j . If you think that the two symbols are difficult to distinguish visually, or at least that there are better choices, go back to page 14 and read footnote 8.

The system is in equilibrium if $d_\xi j = 0$ and $d_\xi \iota = 0$, which is, not surprisingly, the case for $(j, \iota) = (0, 0)$ and $(\rho - \beta/\alpha, 0)$. A linear stability analysis will tell us about the type of these fixed points.¹⁴ The Jacobi matrix of the system is given by

$$J(j, \iota) = \begin{pmatrix} \frac{\partial}{\partial j} (d_\xi j) & \frac{\partial}{\partial \iota} (d_\xi j) \\ \frac{\partial}{\partial j} (d_\xi \iota) & \frac{\partial}{\partial \iota} (d_\xi \iota) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{2\alpha}{D_0} j - \frac{\alpha\rho - \beta}{D_0} & -\frac{c}{D_0} \end{pmatrix}$$

and its eigenvalues at the trivial fixed point $(0, 0)$ are obtained from

$$\det(J(0, 0) - \lambda \mathbb{I}) = \det \begin{pmatrix} -\lambda & 1 \\ -\frac{\alpha\rho - \beta}{D_0} & -\frac{c}{D_0} - \lambda \end{pmatrix} = \lambda^2 + \frac{c}{D_0} \lambda + \frac{\alpha\rho - \beta}{D_0},$$

which has the solutions

$$\lambda^{(0,0)} = -\frac{c}{2D_0} \pm \frac{\sqrt{c^2 - 4D_0(\alpha\rho - \beta)}}{2D_0}.$$

For the other fixed point $(\rho - \beta/\alpha, 0)$ we find the eigenvalues in

$$\det(J(j^*, 0) - \lambda \mathbb{I}) = \det \begin{pmatrix} -\lambda & 1 \\ \frac{\alpha\rho - \beta}{D_0} & -\frac{c}{D_0} - \lambda \end{pmatrix} = \lambda^2 + \frac{c}{D_0} \lambda - \frac{\alpha\rho - \beta}{D_0},$$

¹⁴I do not want to dwell too much on linear stability analysis—here is a really quick and dirty, figuratively speaking overview: The Jacobi matrix $J(j, \iota)$ is the derivative of a multi-dimensional function, here $\mathbb{R}^2 \rightarrow \mathbb{R}^2$, $(j, \iota) \mapsto (d_\xi j, d_\xi \iota)$. Thus, evaluated at a certain point (j, ι) , the Jacobian simultaneously indicates the slope of the two functions $d_\xi j$ and $d_\xi \iota$ for all directions $(\delta j, \delta \iota)$. Right-multiplication with a vector $(\delta j, \delta \iota)^\top$ yields the vector $(d_\xi j, d_\xi \iota)^\top$ indicating the change of j and ι along that direction.

To obtain a picture of how j and ι evolve in the vicinity of a fixed point, we evaluate the Jacobian matrix at that point, $J(j^*, \iota^*)$, and calculate its eigenvalues. Remember that, if λ is an eigenvalue of matrix \mathbf{A} and \mathbf{v} is the corresponding eigenvector, the following relation holds: $\mathbf{A} \mathbf{v} = \lambda \mathbf{v}$.

Thus, for $\mathbf{A} = J(j^*, \iota^*)$ and $\mathbf{v} = (\delta j, \delta \iota)$ a small perturbation around the fixed point, it is clear that along this direction \mathbf{v} the perturbation will grow or shrink proportional to $e^{\lambda \xi}$. A positive eigenvalue λ indicates that the fixed point is unstable with respect to this direction, since the system will move away from it, while a negative λ indicates an stable, attracting fixed point (again, only with respect to this particular direction—since there are two eigenvalues a fixed point may be stable with respect to one eigenvector and unstable with respect to another, resulting in a so-called saddle point). Also, it will be intuitive to the reader that a complex eigenvalue $\lambda = \lambda_r + i\lambda_i$ will somehow introduce some sort of oscillatory behavior due to the factor $e^{i\lambda_i \xi}$.

For a more detailed and rigorous explanation I refer to the excellent book of Strogatz [Str94], which is not only well-written and easy to understand, but also so entertaining that, in theory, the interested and slightly autistic reader may read it solely for the purpose of recreation. Though, admittedly, its price is about one order of magnitude larger compared to that of a novel—probably super-linear with respect to their entertainment values.

$$\lambda^{(j^*, 0)} = -\frac{c}{2D_0} \pm \frac{\sqrt{c^2 + 4D_0(\alpha\rho - \beta)}}{2D_0}.$$

The eigenvalues of the latter, non-trivial, fixed point are always real¹⁵ while the eigenvalues of the former are real only for

$$|c| \geq 2\sqrt{D_0(\alpha\rho - \beta)} = c_{\min}.$$

This is an important result. Imaginary eigenvalues imply that the fixed point is oscillatory, i.e., if $c < c_{\min}$ the solution $j(\xi)$ will converge to the fixed point $j = 0$ in a damped oscillation around zero—which implies that j will be negative for some ξ . This should not be the case since j is a population density and, luckily, is impossible by construction of the model. Thus, we can conclude that if a physically meaningful travelling wave emerges from any initial condition in the system, its velocity will be at least c_{\min} . In fact, according to Murray [Mur89], Kolmogoroff, Petrovsky, and Piscounoff have proven that for any initial condition that has compact support and is zero for large positive x and equal to j^* for large negative x , a travelling wave with $c = c_{\min}$ will emerge, which I faithfully assume to be correct.¹⁶

Up to now we have looked at spatially extended systems with local SIS dynamics, but what about SIR-like diseases? Obviously, *if* the system exhibits travelling waves as well, its form will be different. Since in the SIR model the number of infecteds falls back to zero after some time, we would expect $\lim_{\xi \rightarrow \pm\infty} j(\xi) = 0$. We can make a good argument that travelling waves exist in the SIR model as well, since just as in the SIS model, with a small number of individuals at some location in the system, there will be an exponential increase of infecteds due to the reaction $S + I \rightarrow 2I$ and diffusion will transport them towards neighboring locations. Thus, wave solutions exist—but what about their speed? One might tend to expect a slightly slower wave front because the lacking $I \rightarrow S$ reaction slows down the local reaction process after some time and thus less infecteds are provided to be transported by the diffusion. But watch this: The only time where we explicitly used the SIS model in the above consideration was when we utilized the conservation law $\rho = s + j$ to replace s in the equation. In the SIR model,

¹⁵Except for $\beta \gg \alpha\rho$, but that is not really interesting anyway.

¹⁶Kids, don't try this at home! This is evil Cargo Cult Science [Fey74]. But honestly, I don't really care about the details of this proof (nor about its correctness) because in all my simulations the wave front speed came out to about c_{\min} for the initial condition that I used (see Sec. 3.2). Since my initial condition is slightly different, the proof does not help anyway and the simulation results have to be seen as empirical evidence to the claim that, overlooking their numerical errors, these simulations yield travelling waves with $c = c_{\min}$.

the conservation law reads $\rho = s + j + r$, but initially, there exist no recovered persons and it takes some time until the infected people recover. Thus, the onset of the wave will always be free of recovered persons since the wave will always travel into regions of fully susceptible populations. And it is the onset of the wave which is responsible for the wave propagation since only there the diffusion pushes infected people into new areas. Therefore, since in the onset region $r \approx 0$ and thus $\rho \approx s + j$, the recovered have no significant influence on the wave speed. In fact, we can harden this argument by pointing out that the fixed point in the SIS model which linear stability analysis gave rise to the lower bound on the wave front speed is the same fixed point that exists in the SIR model, namely $(s, j, r) = (1, 0, 0)$ ($s = 1$ was implicitly assumed above but has to be specified in the SIR model to distinguish between the two fixed points $j^* = 0$ before and after the local disease outbreak). Thus, the analysis of the SIS model holds equally well for the SIR model and the wave front speeds in both models are the same.

2.6 Populations that Realize: Different Response Scenarios

In the following, I will introduce three different models that describe the dynamics of a spreading disease in a population that is aware of the forthcoming epidemic wave and reacts to it. In each of the models, a phenomenological approach is taken, i.e., I do not explicitly model every individual's reaction but motivate macroscopic descriptions which can be interpreted to be emerging from a certain microscopic behavior. These macroscopic descriptions are introduced in the framework of spatially continuous mean-field model derived in the previous section.

2.6.1 Panic Reaction: Additional Diffusion

The first model uses a position-dependent diffusion coefficient that is proportional to the density of infected people,

$$D(x) = D_0 (1 + \nu j(x)).$$

Here, ν is a parameter controlling the response intensity. The dynamics of a disease will then be described by the partial differential equation already derived in Sec. 2.5 (R denoting the local disease dynamics term),

$$\partial_t j = R + \partial_{xx}(D j),$$

in particular,

$$\partial_t j = R + D_0 \partial_{xx} j + 2 D_0 \nu \left(j \partial_{xx} j + (\partial_x j)^2 \right).$$

The locally increased diffusion can be interpreted as an undirected panic reaction: the more infecteds are at a person's location, the higher the probability that this person will run away—in whatever direction.

2.6.2 Directed Flight: Gradient Force

In contrast to the undirected panic flight, it would seem more appropriate to model something like a directed flight where people run away from the forthcoming wave. One way to create such a model is to introduce a drift that is proportional to the negative spatial derivative of the density of infecteds,

$$F(x) = -\mu \partial_x j.$$

Thus, people will have a higher probability of travelling towards regions with lower density of infecteds. The parameter μ is used to control the strength of the population's response. The dynamics of an SIS-like disease with directed flight reaction will then be governed by

$$\partial_t j = \alpha s j - \beta j + D_0 \partial_{xx} j - \mu \partial_x (j \partial_x j).$$

2.6.3 Strategic Flight: Integral Force

The third strategy I want to investigate is similar to the previous one in the regard that both add a drift term to the dynamical equation. However, the force in the directed flight reaction was local and thus responses were restricted to an already arrived disease. Here, I will introduce an alternative drift function that models a more foreseeing, strategic flight reaction: People will already flee when a forthcoming epidemic wave is still some distance away. To implement this, we express the drift similarly to a gravitational force with the infecteds serving as a *repelling* mass:

$$F(x) = -\mu \int_{-\infty}^{\infty} \frac{x-y}{|x-y|} K(|x-y|) j(y) dy.$$

The integration kernel $K(d)$ must of course be a function that quickly goes to zero as $d \rightarrow \infty$. It determines at which distances the mass of infected people will



Figure 2.7: Integration kernel used in the strategic flight scenario.

have stronger or weaker repelling effect. If $K(d)$ is a Gaussian with mean zero, $K(d) = A \exp(-d^2/(2\sigma^2))$, the effect is somewhat similar to the gradient force: People are forced into the direction where less infecteds are and the force is stronger if the difference between the number of infecteds on both sides is greater. However, if the kernel is a function that is small around zero and raises to a peak at $d = d_{\max}$, it will model a special kind of behavior: People will flee from a forthcoming wave if it is sufficiently far away, but if the wave is too near the kernel reduces the repelling effect of the infecteds and people will not flee anymore—they give up and stay in the infected area. In this study, I use a sum of two Gaussians,

$$K(d) = Z \left(\exp\left(-\frac{(d - d_{\max})^2}{2\sigma^2}\right) + \exp\left(-\frac{(d + d_{\max})^2}{2\sigma^2}\right) \right),$$

where d_{\max} and σ are parameters and Z is an appropriate normalization constant, such that $\int_0^\infty K(x) dx = 1$.

2.6.4 Corresponding Strategies in Stochastic Models

In order to compare the results of simulated spatially continuous systems with stochastic simulations, we need to translate the models motivated above into stochastic formulations. From the derivation of the Fokker-Planck equation in Sec. 2.5 we can see that there is a general rule to create a corresponding stochastic model. For any given diffusion and drift functions $D(x)$ and $F(x)$ (which may be time-dependent, of course), we have to discretize them into coefficients $D_i = D(x_0 + i \Delta x)$ and $F_i = F(x_0 + i \Delta x)$ and then define travel rates as

$$\begin{aligned} w_i^+ &= D_i + |F_i| \delta_{\text{sgn}F_i, 1} = D_i + \check{\Theta}(F_i), \\ w_i^- &= D_i + |F_i| \delta_{\text{sgn}F_i, -1} = D_i + \check{\Theta}(-F_i). \end{aligned}$$

Put in words, this means that the travel rates are both set to the diffusion coefficient and then the drift term is added to the travel rate into the right (left) city if the drift coefficient is positive (negative). For notational simplicity I define a modified Heaviside function

$$\check{\Theta}(x) := |x| \Theta(x) = \begin{cases} x & x > 0 \\ 0 & \text{else} \end{cases}.$$

Applied to the three response strategies this implies that the panic reaction is modeled by

$$w_i^+ = w_i^- = D_0 (1 + \nu j_i).$$

The directed flight strategy, where $F(x) = -\mu \partial_x j$, translates into

$$\begin{aligned} w_i^+ &= D_0 + \check{\Theta}\left(-\frac{\mu}{\Delta x} (j_{i+1} - j_i)\right), \\ w_i^- &= D_0 + \check{\Theta}\left(-\frac{\mu}{\Delta x} (j_{i-1} - j_i)\right). \end{aligned}$$

Note that I used the forward discretization of the first derivative for the travel rate into the right city and the backward discretization in the other case—the more intuitive approach would probably have been to use the centered difference $(j_{i+1} - j_{i-1})/(2\Delta x)$ in both cases. But this has an interesting advantage: The observing reader may note that the two equations look alike and can be merged into one single equation describing the travel rate from city j into city i :

$$w_{ij} = D_0 + \check{\Theta}\left(-\frac{\mu}{\Delta x} (j_j - j_i)\right).$$

Thus, the travel rates are formally independent of the geographical order of the nodes. If we set $\Delta x = 1$, the geographical notion completely disappears from the equation and we have generalized the directed flight strategy to arbitrary networks—though, of course, only in the case of a linear chain it would be the stochastic equivalent of the spatially continuous model introduced above.

The strategic flight behavior can also be translated into stochastic terms and the interested reader will be able to easily figure this out, but since no comparative simulations have been performed by Rafael Brune for his diploma thesis [Bru08], I will omit it.

3 Methods

This chapter is divided into three parts. I will first present some general concepts for numerically solving partial differential equations that exhibit travelling wave solutions. Afterwards, I give the methodological details used in my study. The last section investigates some numerical details of the implementation which are not so important for understanding and interpreting the results.¹ In the next chapter, I also compare my results to those obtained by Rafael Brune in his stochastic simulations. For a derivation and explanation of the utilized methods I refer to his diploma thesis [Bru08].

3.1 Concepts

3.1.1 Solving Ordinary Differential Equations

Ordinary differential equations are of the form

$$d_t x = f(x, t)$$

and their solution is obtained by integration

$$x(t) = x(t_0) + \int_{t_0}^t f(x, t') dt'.$$

Unfortunately, some functions $f(x, t)$ refuse to be analytically integrated and here numerical integration comes into play. The basic idea is to discretize the differential equation to

$$x(t + \Delta t) = x(t) + f(x(t), t) \Delta t$$

and obtain the solution for a given initial condition $x(t_0)$ by subsequently summing the pieces,

$$x(t_n = t_0 + n \Delta t) = x(t_0) + \sum_{i=0}^{n-1} f(x_i, t_0 + i \Delta t) \Delta t.$$

¹Here, “not so important” is a euphemism for “completely irrelevant,” but the investigation yields some interesting insights and therefore I included it anyway.

Of course, this so-called *Euler method* gives only an approximation (it linearly extrapolates the slope of x at time t over the whole interval Δt), but for $\Delta t \rightarrow 0$ the solution obtained by numerical integration converges to that of the analytical integration. Thus, for sufficiently small Δt , the numerical solution is a good approximation. Note, however, that a smaller Δt implies more computational effort, i.e., longer computing time, since more terms in the sum have to be calculated.

How good the approximation is can be seen when doing a Taylor series expansion of $x(t)$. We expand around t and our small deviation from t will be Δt . Then, the Taylor expansion reads

$$x(t + \Delta t) = \sum_{n=0}^{\infty} \frac{1}{n!} d_t^n x(t) \Delta t^n = x(t) + d_t x(t) \Delta t + \frac{1}{2} d_t^2 x(t) (\Delta t)^2 + \dots$$

Here, $d_t^n x(t)$ reads “the n -th derivative of x evaluated at t ” which is a scalar value (remember that t is a fixed value here and that Δt is the independent variable of the expansion). Now, this was easy: The first two terms are identical to the discretized differential equation of the Euler scheme. Thus, in each step, the Euler scheme drops terms of $\mathcal{O}(\Delta t^2)$ and thus is only accurate to first order in Δt .

We would improve accuracy to second order if we were able to calculate the second derivative $d_t^2 x(t)$, which is $d_t f(x, t)$. This we can write as

$$d_t f(x, t) = \partial_x f d_t x + \partial_t f = f \partial_x f + \partial_t f.$$

Unfortunately, the function $f(x, t)$ might be so nasty that it not only refuses to be analytically integrated but also does not want to be differentiated—not even partially. But we might be able to find an approximation. The (two-dimensional) Taylor expansion of f reads

$$f(x + \xi, t + \tau) = f(x, t) + \xi \partial_x f(x, t) + \tau \partial_t f(x, t) + \dots$$

With $\xi = \frac{1}{2} f(x, t) \Delta t$ and $\tau = \frac{1}{2} \Delta t$ we get, to first order,

$$f(x + \frac{1}{2} f(x, t) \Delta t, t + \frac{1}{2} \Delta t) \approx f(x, t) + \frac{1}{2} (f(x, t) \partial_x f(x, t) + \partial_t f(x, t)) \Delta t.$$

Note the $d_t f(x, t)$ hiding in the second term. What we have here is the derivative of $x(t)$ at the midpoint between the times t and $t + \Delta t$ where the value $x(t + \frac{1}{2} \Delta t)$ is approximated by an Euler step. Replacing $f(x, t)$ in the Euler step with this midpoint derivative yields an update rule proposed by Carl Runge in his interesting 1895 paper [Run95] and also known as the *midpoint method*,

$$\begin{aligned} x(t + \Delta t) &= x(t) + f(x + \frac{1}{2} f(x, t) \Delta t, t + \frac{1}{2} \Delta t) \Delta t \\ &= x(t) + f(x, t) \Delta t + \frac{1}{2} d_t f(x, t) (\Delta t)^2. \end{aligned}$$

Since $d_t x = f(x, t)$, this update rule agrees with the Taylor expansion of $x(t)$ up to second order.

By choosing alternative values for ξ and τ , namely $\xi = f(x, t) \Delta t$ and $\tau = \Delta t$, we obtain, to first order, an approximation of $d_t f(x, t)$:

$$\frac{f(x + f(x, t) \Delta t, t + \Delta t) - f(x, t)}{\Delta t} \approx f(x, t) \partial_x f(x, t) + \partial_t f(x, t).$$

This we can plug into the Taylor expansion of $x(t)$ yielding another second-order update rule also proposed by Runge in the same paper, but known as the *Heun method*, named after Karl Heun who published it again in 1900 [Heu00],

$$\begin{aligned} x(t + \Delta t) &= x(t) + f(x(t), t) \Delta t + \frac{1}{2} (f(x + f(x(t), t) \Delta t, t + \Delta t) - f(x(t), t)) \Delta t \\ &= x(t) + \frac{1}{2} (f(x + f(x(t), t) \Delta t, t + \Delta t) + f(x(t), t)) \Delta t. \end{aligned}$$

The basic idea here is to do an Euler step across the whole interval Δt , then evaluate the derivative of $x(t)$ at the end of the interval and redo the update step using the mean of the derivatives at t and $t + \Delta t$.

Thus, we have two mechanisms to improve accuracy: One is to use the slope at the midpoint of the interval, which requires an estimation of x at the midpoint, and the other is to incorporate a second slope into the linear extrapolation. Intuitively, it should be possible to add more slopes into the extrapolation or include differently estimated midpoint slopes to further improve accuracy. In his 1901 paper [Kut01], Martin Kutta systematically generated a class of discrete update rules that use different combinations of the two mechanisms to reproduce more and more terms of the Taylor expansion of $x(t)$. This class of update rules is now known as the *Runge-Kutta methods* and the most famous is the following, accurate to fourth

order:

$$\begin{aligned} s_1 &= f(x, t), \\ s_2 &= f(x + s_1 \Delta t/2, t + \Delta t/2), \\ s_3 &= f(x + s_2 \Delta t/2, t + \Delta t/2), \\ s_4 &= f(x + s_3 \Delta t, t + \Delta t), \\ x(t + \Delta t) &= x(t) + \frac{1}{6} (s_1 + 2s_2 + 2s_3 + s_4) \Delta t. \end{aligned}$$

The first, s_1 , is the slope at the start of the interval, i.e., the slope used in the Euler method. The next two are both slopes at the midpoint of the interval, but s_3 is calculated with $x(t + \frac{1}{2} \Delta t)$ estimated by the slope s_2 , i.e., similar to the Runge method. Finally, s_4 is the slope at the end of the interval as in the Heun method and the slope used for the final update step is an average of the four derivatives with an emphasis on the two midpoints. According to most lecture notes introducing this classical Runge-Kutta method, the proof of it being accurate to fourth order is “beyond the scope of this document” and “simple but tedious,” involving the two-dimensional Taylor expansion of $f(x, t)$ up to third order which produces the terms of the expansion of $x(t)$ up to fourth order. But they always point out that it can be found in “advanced texts.” Thus, I will simply trust the large number of scientists who have already faithfully used this method.²

3.1.2 Solving Partial Differential Equations

With the methods described above we can solve ordinary differential equations but the systems I want to study are always described by partial differential equations involving both temporal and spatial derivatives. However, the problem of solving partial differential equations can be reduced to the problem of solving ordinary differential equations: We revert the process of going to the continuum limit made in Sec. 2.5 and discretize the space axis into lattice sites with locations $x_i = x_0 + i \Delta x$. Now, the density of infected people $j(x, t)$ is again represented by a discrete set of functions $j_i(t)$ which form a set of coupled ordinary differential equations solvable by the methods described above.

In practice, how is this spatial discretization achieved? There are three possibilities: the forward difference

$$\partial_x j \approx \frac{j(x + \Delta x) - j(x)}{\Delta x},$$

²Sorry, Feynman.

the backward difference

$$\partial_x j \approx \frac{j(x) - j(x - \Delta x)}{\Delta x},$$

and the centered difference

$$\partial_x j \approx \frac{j(x + \Delta x) - j(x - \Delta x)}{2 \Delta x}.$$

Each of them converges to the exact derivative in the limit $\Delta x \rightarrow 0$. The first two (asymmetrical) approximations are exact to first order, as the observing reader will quickly figure out by looking at the (one-dimensional) Taylor expansion of $j(x, t)$ in x for fixed time t . The centered difference, as the average of the two first approximations, is accurate to second order since in the Taylor expansion the second order terms cancel each other:

$$\begin{aligned} \frac{j(x + \Delta x) - j(x - \Delta x)}{2 \Delta x} &= \frac{1}{2 \Delta x} (j(x) + \partial_x j(x) \Delta x + \frac{1}{2} \partial_{xx} j(x) (\Delta x)^2 + \dots \\ &\quad - j(x) + \partial_x j(x) \Delta x - \frac{1}{2} \partial_{xx} j(x) (\Delta x)^2 + \dots) \\ &= \partial_x j(x). \end{aligned}$$

Thus, the centered difference is the more accurate choice for discretizing the partial derivative.³ Of course, in addition to the initial condition $j(x_i, t_0)$ which provides initial values for j at each site, we also have to specify *boundary conditions* $j(x_0 - \Delta x, t)$ and $j(x_N + \Delta x, t)$ for all times t to calculate the centered difference at $x = x_0$ and $x = x_N$ (where $N + 1$ is the number of lattice sites used in the discretization).

The diffusion term of the equations introduced in Sec. 2.5 also involves the second spatial derivative. Of course, we could use the centered difference of centered

³The observing reader may object that we could use the backward difference for the time derivative as well (instead of the forward difference used in the Euler and Runge-Kutta methods), or that we should use the centered difference. The problem is, that, in the notation of the last section, we then gain equations like

$$x(t) = x(t - \Delta t) + f(x(t), t) \Delta t,$$

where the variable of interest, $x(t)$, is *implicitly* defined (it appears on both sides of the equation, thus we would have to solve the equation for x), or

$$x(t + \Delta t) = x(t - \Delta t) + 2 f(x(t), t) \Delta t,$$

for which we would need two initial conditions $x(t_0)$ and $x(t_0 - \Delta t)$. That might not seem to be a problem, but while we can choose $x(t_0)$ freely to select one of trajectories that are solutions of the equation, the second initial condition $x(t_0 - \Delta t)$ has to be a point on the same trajectory (which is unknown a priori). Thus, unless you have a well-behaved system, where the backward difference, leading to the implicit update rule, might be advantageous, it is best to stick to the forward difference, yielding a straightforward *explicit* update rule.

differences around $j(x - \Delta x)$ and $j(x + \Delta x)$. But then, the calculation of the second derivative involves the values of j at five sites: $j(x)$, $j(x \pm \Delta x)$, and $j(x \pm 2 \Delta x)$. For this we would need additional boundary conditions that extend two sites beyond the lattice instead of only one. Instead, we can use the backward difference of forward differences,

$$\begin{aligned} \partial_{xx} j &\approx \frac{\partial_x j(x) - \partial_x j(x - \Delta x)}{\Delta x} \approx \frac{\frac{j(x + \Delta x) - j(x)}{\Delta x} - \frac{j(x) - j(x - \Delta x)}{\Delta x}}{\Delta x} \\ &= \frac{j(x + \Delta x) - 2j(x) + j(x - \Delta x)}{(\Delta x)^2}, \end{aligned}$$

to obtain a compact symmetrical second derivative which is accurate to second order:

$$\begin{aligned} &\frac{j(x + \Delta x) - 2j(x) + j(x - \Delta x)}{(\Delta x)^2} \\ &= \frac{1}{(\Delta x)^2} (j(x) + \partial_x j(x) \Delta x + \frac{1}{2} \partial_{xx} j(x) (\Delta x)^2 + \dots \\ &\quad - 2j(x) + j(x) - \partial_x j(x) \Delta x + \frac{1}{2} \partial_{xx} j(x) (\Delta x)^2 + \dots) \\ &= \partial_{xx} j(x). \end{aligned}$$

It seems that we have all the necessary tools for solving partial differential equations now, but there is one thing still to consider: The two discretizations Δt and Δx cannot be freely chosen, as we will see in the simple diffusion equation

$$\partial_t j = D_0 \partial_{xx} j.$$

The solution $j(x, t)$ can be written as a complex Fourier series, i.e., we compose the function by using time-dependent coefficients to spatially oscillating exponentials,

$$j(x, t) = \sum_{k=-\infty}^{\infty} \hat{j}_k(t) e^{i k x}.$$

Then, discretizing the diffusion equation yields

$$j(x, t + \Delta t) = j(x, t) + D_0 \frac{j(x + \Delta x, t) - 2j(x, t) + j(x - \Delta x, t)}{(\Delta x)^2} \Delta t,$$

$$\hat{j}_k(t + \Delta t) = \hat{j}_k(t) \left(1 + \frac{D_0 \Delta t}{(\Delta x)^2} (e^{i k \Delta x} - 2 + e^{-i k \Delta x}) \right).$$

The second equation states the discretized update rule for the Fourier coefficients, obtained by plugging the Fourier series into the first equation and then individually equating the terms for each mode k . We see that the mode amplitude \hat{j}_k is amplified by a factor

$$1 + \frac{D_0 \Delta t}{(\Delta x)^2} (e^{ik\Delta x} - 2 + e^{-ik\Delta x}) = 1 + 2 \frac{D_0 \Delta t}{(\Delta x)^2} (\cos(k\Delta x) - 1).$$

If this factor's absolute value is greater than 1, the corresponding mode will be amplified in *every* time-step and the simulated system quickly explodes into a sparkling cloud of nan and inf values. How can we prevent such a disaster? If $\cos(k\Delta x) = 1$ the factor will be 1—and there is no possibility for it to be larger than unity. Thus, the worst case is $\cos(k\Delta x) = -1$ and we wish the factor to be larger than -1 in this case, i.e.,

$$1 - 4 \frac{D_0 \Delta t}{(\Delta x)^2} > -1 \quad \Leftrightarrow \quad \Delta t < \frac{(\Delta x)^2}{2D_0}.$$

This inequality imposes a restriction on how small we have to choose Δt for a given Δx if we want a numerically stable solution. The presented method is thus named *von Neumann stability analysis*—after John von Neumann, who developed it for Los Alamos' Manhattan.

3.1.3 Following Travelling Waves with a Moving Simulation Window

The systems under investigation in this study exhibit travelling waves. As shown in Sec. 2.5.1, the wave form $j(\xi)$ approaches fixed points for $\xi \rightarrow \pm\infty$. Thus, if we discretize the system into an infinite lattice, most sites will have values s_i , j_i and r_i very close to either of the fixed points. The real action takes place only in a small confined area $x_{\min} < x < x_{\max}$, in which the wave front currently is. Outside this area, the change of j_i etc. is essentially zero. Thus, instead of simulating a large lattice over which the wave travels, it should be sufficient to simulate only a small number of lattice sites around the wave front⁴ and move the simulation area as the wave travels towards one end, which is realized by removing one lattice site from the left end of the system and adding one lattice site to the right end (assuming the wave travels from left to right). This technique allows the observation of the wave for arbitrary long times.

⁴Of course, the lattice still has to be large enough to cover the complete area that influences the dynamics. A detailed investigation on this will be presented in Sec. 3.3.

3.2 Details

This section briefly provides various details on how I applied the methodological concepts to the theoretical framework derived in Chapter 2.

Following the considerations from the last section, it seems appropriate to use the classical fourth-order Runge-Kutta method for integrating the system of coupled ordinary differential equations obtained by the spatial discretization for which, of course, the centered difference equations were used.

As initial condition, a completely susceptible and homogeneously distributed society was implemented and a small amount of infected people placed into the left-most site of the simulation area: $j(x_i, t = 0) = j_0 \delta_{i,0}$, $s(x_i, 0) = 1 - j(x_i, 0)$, $r(x_i, 0) = 0$. From this initial condition, a stable wave front emerges after some equilibration time. To follow the wave front, the moving simulation window algorithm described above was implemented. Initially, the simulation area covered the space $x \in [x_{\min}, x_{\max}]$ with $x_{\min} < 0 < x_{\max}$. The site with $x = 0$ at $t = 0$ was used as an anchor point for the simulation window. When using a model with a non-zero fixed point (SIS), the simulation window was shifted by one lattice site if the density of infecteds at the anchor point exceeded half of the fixed point value. With models in which the density of infecteds is zero on both sides of the wave (SIR), the window was moved if the maximum of infecteds was at a site with $x > 0$. In each window shift, a fully susceptible lattice site with total population density $\rho = 1$, i.e., identical to the initial condition, was added to the right.

Regarding the boundary condition, I assumed $j(x_N + \Delta x, t) = j(x_N, t = 0)$, and analogously for s and r . In other words, beyond the right end of the simulation area, a fully susceptible site was assumed for all times t , which is consistent with the moving window implementation. To the left, I assumed $j(x_0 - \Delta x, t) = j(x_0, t)$, i.e., the first site beyond the left end was assumed to be identical to the left-most site within the simulation area. This is consistent with the fact that the wave form $j(\xi)$ converges to a constant value for $\xi \rightarrow -\infty$.

The implementation of the additional diffusion in the panic reaction scenario and of the gradient force in the directed flight scenario is straightforward—of course using the centered difference for calculating the gradient force. For the calculation of the integral force in the strategic flight scenario some simplifications have been made, since computing the complete integral for every evaluation of the right hand side of the differential equations is computationally extensive (the number of operations scales with N^2 , since for each of the N sites the integral has one term for each site). The most intuitive way to compute the integral force at some place

would be to iterate for each lattice site i over a small neighborhood $j \in [i-m, i+m]$ and add up all the terms for each site j to the integral value at site i . However, in the implementation I reversed the order of the two iterations: For each lattice site j , terms were added to each site $i \in [j-m, j+m]$ if, and this is the simplification, the density of infecteds at site j , $j(x_j, t)$, was above a certain threshold $\theta_{\text{int}} = 10^{-5}$. This way, a lot of low-impact terms could be skipped. Also, the calculation of the integral force was performed only once for each of the Runge-Kutta steps and not for each of the four sub-steps, i.e., not for each individual evaluation of $f(x, t)$. Finally, in the two-dimensional simulations, the kernel function $K(d)$ was precomputed for integer values of d and later the kernel values for arbitrary real values of d were approximated by the precomputed value of the largest smaller integer, $\lfloor d \rfloor$. By using a precomputed lookup table, the time-intensive calculation of two exponential function values per integral term could be avoided. Furthermore, the lookup table was truncated at some value m for which $K(d)$ was negligibly small (i.e., for which the integral term would not add any significant value to the total integral). This also led to a significant reduction of the number of integral terms to be calculated.

For measuring the wave front speed and form with different underlying models, response scenarios and response strengths, I set up one system per model and scenario. After some initial equilibration time with response strength zero, I measured the wave front velocity by stopping the time needed for a predefined number of window shifts. Then, iteratively until the full parameter range of interest was covered, the response strength parameter was increased by a small amount and the velocity was measured again after another short equilibration time. After each velocity measurement the (equilibrated) wave form was recorded.

For obtaining the phase diagram of the strategic flight scenario, I set up a system $x \in [0, 100]$ and triggered an outbreak at $x = 0$. To prevent an arbitrary small amount of infected people leaking through an evacuated low-density region and causing an outbreak on the other side, an infection threshold θ was used, as explained in the next paragraph. The simulation was carried out until either the simulation time exceeded $t_{\text{max}} = 50$ or at each lattice site the density of infecteds dropped below the infection threshold, in which case the disease was considered to be extinct. I measured the maximum total population density and the maximum x which had infected people during the simulation. If the density of infecteds at $x = x_{\text{min}} + (x_{\text{max}} - x_{\text{min}})/3$ exceeded the threshold value θ , the simulation window was moved.

In the density formulation of the models (as opposed to absolute numbers of

susceptibles, infecteds and recovered) very small densities occur, which already have a significant impact on the dynamics—even the smallest non-zero density of infecteds triggers an exponential growth of j . However, in the stochastic system, a population cannot be infected unless at least one infected individual is brought into the population. We can capture this in the density formulation by introducing a threshold θ and insisting on $j \geq \theta$ to allow infections. The differential equation for the density of infecteds in the SIS system with threshold reads

$$d_t j = \alpha s j \Theta(j - \theta) - \beta j + \partial_{xx}(D j) - \partial_x(F j).$$

Of course, the factor $\Theta(j - \theta)$, which is 0 for $j < \theta$ and 1 otherwise, also appears in the corresponding term in the equation for $d_t s$. The threshold will have a significant effect not only on the outbreak dynamics as mentioned above, but also on the equilibrium behavior of the system, as we will see in the next chapter.

Finally, I introduced another parameter to play with. In the equations for $d_t j$ and $d_t r$, the response strength parameters ν and μ were replaced by $\nu \delta_I$ and $\mu \delta_I$ where δ_I is a binary switch (either 0 or 1) stating whether the flight behavior applies to infecteds as well. Thus, I was able to investigate the difference between the cases where everybody is running away from a forthcoming wave front or where only the susceptible individuals run away. Also, this additional parameter δ_I will turn out to have significant impact.

A few simulations have been performed in two spatial dimensions where $j = j(\vec{x}, t)$ with $\vec{x} = (x, y) \in \mathbb{R}^2$. The mean-field equations for this case arise naturally from the one-dimensional equations by replacing the spatial derivatives with the Nabla operator $\nabla = (\partial_x, \partial_y)$:

$$d_t j = R + \nabla^2(D j) - \nabla(F j).$$

Similarly for $d_t s$ and $d_t r$ and in the definitions for $F(\vec{x}, t)$ in the different response scenarios,

$$F(\vec{x}, t) = -\mu \nabla j, \quad F(\vec{x}, t) = -\mu \iint \frac{\vec{x} - \vec{y}}{\|\vec{x} - \vec{y}\|} K(\|\vec{x} - \vec{y}\|) j(\vec{y}) d\vec{y}.$$

Also, the discretization for the numerical solution arises naturally: The underlying lattice is two-dimensional and the partial derivatives ∂_x and ∂_y are both replaced by centered differences.

Unless otherwise specified, $\alpha = 3$, $\beta = 1$, and $D_0 = 1$ was used in all simulations.

3.3 Details' Details

Choosing the right Δt and Δx can be a difficult decision since it involves a trade-off between computation time and accuracy. In the following, I want to present a systematic investigation on how the discretization details impact the accuracy of the measured wave front velocity. The effects are very subtle, too small to be relevant for the results presented in the next chapter, so you might as well skip this section—unless you are interested in unimportant (but interesting) numerical details.

Figure 3.1 shows simulated trajectories of single SIS and SIR populations with different time steps Δt using the Euler method. The density of infecteds increases superlinearly for small times and thus, in each step the Euler method underestimates the derivative of j in the interval $[t, t + \Delta t]$. This explains the delay in the raise of infecteds and the time at which the window shift criterion is first fulfilled (Fig. 3.1, right). Using the Runge-Kutta method will of course improve accuracy, though qualitatively the effects are the same. Therefore, we would expect that a large Δt leads to a slower wave front velocity in our spatially continuous system, since the infection dynamics run slightly slower.

Next, I investigated the impact of the lattice spacing Δx on the time scales in simple diffusion governed by the equation $\partial_t \rho = \partial_{xx} \rho$. A test area $x \in [-2.5, 2.5]$ was set up and initialized with a discretization independent amount of individuals at the center, $\rho(x_i = 0, t = 0) = 1/\Delta x$, $\rho(x_i \neq 0, t = 0) = 0$. To avoid significant errors from the time discretization, a small time step $\Delta t = 10^{-6}$ was chosen. After large time, the diffusion levels the peaked population distribution to a constant ρ , thus the time where $\rho(x = 1)$ is half of $\rho(x = 0)$ can be interpreted as a measure for the diffusion speed. Figure 3.2 shows the simulation results for various Δx . Since the measure time increases with Δx we would again expect a slower wave front velocity for larger lattice spacings.

This hypothesis was tested using the SI model⁵ in a simulation area of length $L = 150$ with $x_{\min} = -100$ and $x_{\max} = 50$ represented by $N = L/\Delta x$ lattice sites. I equilibrated the system until 10 shifts of the moving simulation window have occurred and then the measurement was carried until the window has moved by one area size L , after which the wave front velocity is $c = L/T$, where T is the time needed for the measurement.

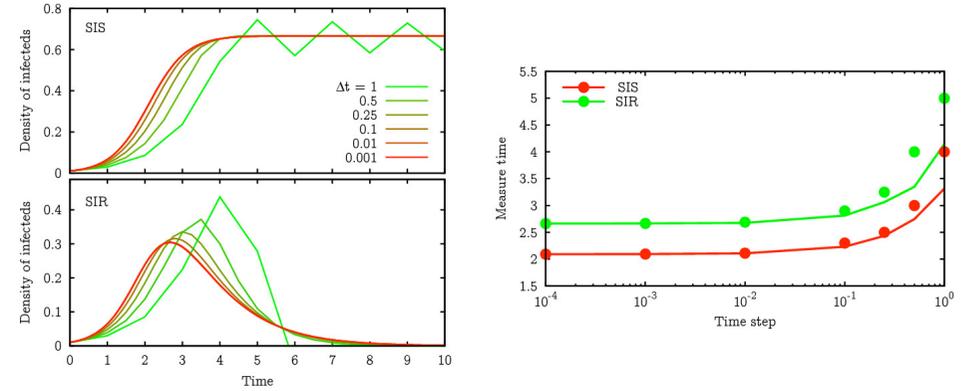


Figure 3.1: Accuracy of SIS and SIR time evolution. The two left plots show the time evolution of the density of infecteds in the SIS (top) and SIR (bottom) model for $\alpha = 3$ and $\beta = 1$ obtained with different time discretizations Δt . In each simulation, the time when the density of infecteds first exceeded half the fixed point density (SIS) or first dropped below the value from the previous time step (SIR) was measured. These times are visualized in the right plot: Dots show the real measured time and lines indicate a refinement obtained by linear interpolation.

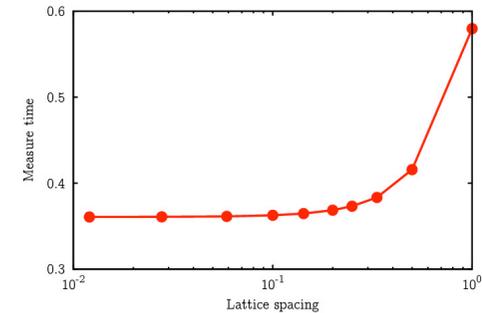


Figure 3.2: Accuracy of simple diffusion. For different spatial discretizations Δx in a system of size $L = 5$ and initial population density distribution $\rho(x, t = 0) = \delta(x)/\Delta x$, the time was measured where first $\rho(x = 1) = \rho(x = 0)/2$.

⁵As indicated in Fig. 2.3 on page 12, the SI model has no recovery reaction $I \rightarrow S$ and therefore runs into a stable fixed point with $j^* = 1$. It is expected to exhibit travelling waves with velocity $c_{\min} = 2\sqrt{\alpha D_0}$.

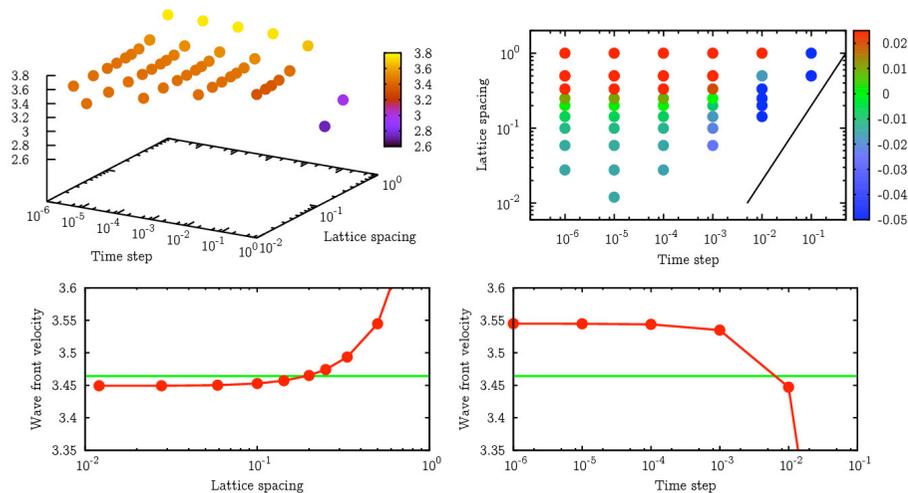


Figure 3.3: Accuracy of the speed measurement: Euler. For the first measurement method described in the text (simulation area with $x \in [-100, 50]$, equilibration for 10 window shifts and measurement for $N = L/\Delta x$ window shifts), the observed wave front velocity in the spatially continuous SI model ($\alpha = 3$, $D_0 = 1$) is shown for different values of the time step Δt and the lattice spacing Δx . The top left figure gives an overview of the speed surface, the top right shows the absolute deviations from the analytical value $c = 2\sqrt{D_0}\alpha$. The diagonal line indicates the stability criterion $2\Delta t < (\Delta x)^2$ for the ordinary diffusion equation (except for the small Δx at $\Delta t = 10^{-6}$, missing points are due to numerical instabilities). In the bottom row, slices from the data are presented for $\Delta t = 10^{-5}$ (left) and $\Delta x = 0.5$ (right), in comparison with the analytical value (green).

Figure 3.3 displays the results, which show that the measured velocity independently converges for $\Delta t \rightarrow 0$ and $\Delta x \rightarrow 0$. The wave appears slower for large Δt which is consistent with the above consideration. However, the velocity is *higher* for larger values of Δx . This can only be explained by the reaction dynamics: Though the diffusion is slower for large Δx , i.e., the site at $x = 1$ will gain less infecteds by diffusion than with a smaller Δx , the exponential growth due to the infection dynamics can outrun this. With a certain amount of infecteds at site $i = 0$, after one time step there will be a small amount of infecteds carried over to site $i = 1$ and the exponential growth can take over. If Δx is large, this neighboring site is further apart and with a ten times smaller Δx it takes nine additional time steps until a non-zero density of infecteds is found at the site with the same x .

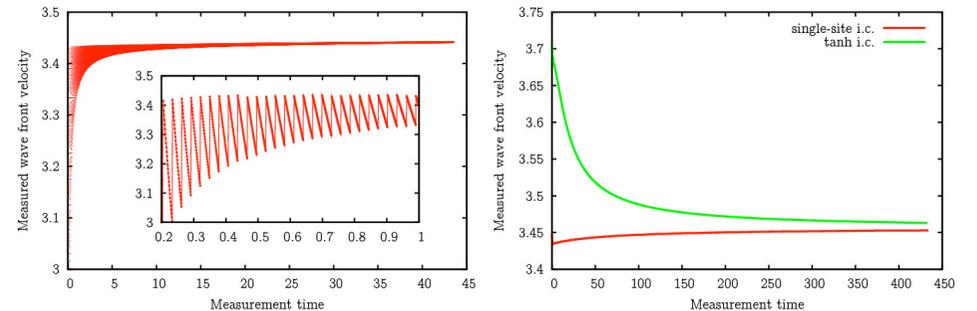


Figure 3.4: Continuous velocity measurements in the SI model. The plots show the velocity $m(t)\Delta x/(t - t_0)$ where $m(t)$ denotes the number of window shifts between t_0 and t .

Also note that the asymptotic value of the wave front velocity is slightly below the analytical value. The relative error is just 0.4%, so the results are absolutely acceptable, but I was interested in understanding from where the error arises. Instead of just measuring the velocity at the end of the simulation, I recorded a velocity $\check{c}(t) = m(t)\Delta x/(t - t_0)$ after every simulation time step, where $m(t)$ denotes the number of window shifts since the first shift at time t_0 .

Figure 3.4 (left) shows $\check{c}(t)$ for $\Delta t = 10^{-3}$ and $\Delta x = 0.1$. Note that the measured velocity $\check{c}(t)$ converges from below to the “real” velocity. The inset illustrates the saw-tooth-like fine structure of the curve. Each jump to a higher velocity corresponds to a simulation window shift. After each shift only the time increases but not the measured distance and thus the velocity decreases until the next window shift occurs. Though the amplitude of this oscillation vanishes for infinite simulation time, we can easily get rid of this effect by allowing velocity measurements only in time steps with window shifts. This is not as arbitrary as it might seem at first sight: At the time step where the window shift occurs we know the exact time when the wave passed the position $x = 0$, where “passed” means “density of infecteds fulfills the window shift condition” and “exact time” means “exact time within simulation accuracy Δt .” Thus, in the time steps with window shifts we have the most exact knowledge of the wave front’s spatio-temporal position.

The right plot in Fig. 3.4 shows the velocity evaluated only at the time steps with window shifts for two different initial conditions: First, the already mentioned

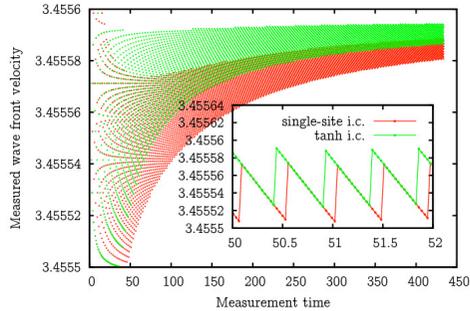


Figure 3.5: Continuous velocity measurements after long equilibration. The setup was the same as for Fig. 3.4, but the equilibration was carried out for the time needed for $10N$ window shifts.

single-site initial condition given by

$$j(x_0, t = 0) = 10^{-2}, \quad j(x_i, t = 0) = 0 \quad (i = 1, 2, 3, \dots),$$

and second, a more equilibrium-like initial condition given by

$$j(x_i, t = 0) = \frac{1}{2} (\tanh(-0.6(x_i - x_{\min}/2)) + 1).$$

One sees from the figure, that, as seen in the left plot, the velocity converges from below for the single-site initial condition while the velocity measured after the tanh initial condition converges from above. Please also note the longer simulation time in the right plot. Although the velocity seems to converge fine in the left plot, we can see in the right one that it significantly varies in the second decimal place—and even in the first decimal digit with the tanh initial condition.

Apparently, the system is not in equilibrium yet. Therefore, I allowed for a longer equilibration delaying the measurement until the simulation window moved across an area ten times the lattice size L (instead of only 10 single window shifts). The results are shown in Fig. 3.5. Both initial conditions now have converged to nearly the same value, differing in the fifth decimal place only. Interestingly, one observes the same saw-tooth-like structure of the curve (remember that here I am plotting only the local maxima of the saw-tooth structure seen in Fig. 3.4), which is an undersampling effect: Due to the finite update time, the condition for the window shift is not evaluated at every time t but only at discrete times t_n and thus the statement “window shift condition satisfied at time t_n ” is always to be read as

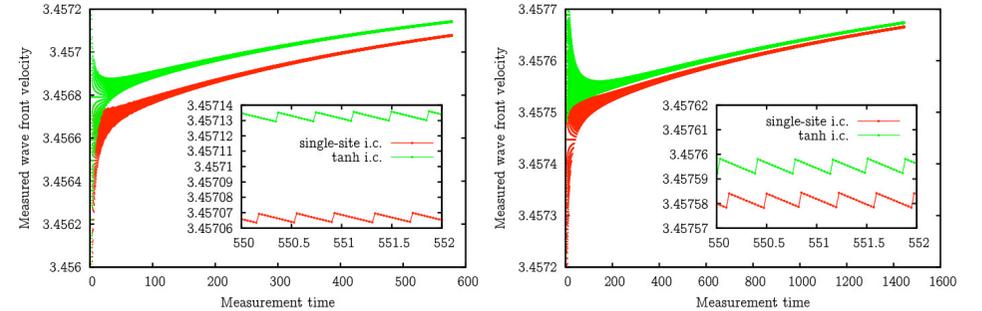


Figure 3.6: Continuous velocity measurements in large simulation areas. Again, a long equilibration was done before starting the measurement in a simulation area with $x \in [-100, 100]$ (left) and $x \in [-250, 250]$ (right).

“window shift condition was satisfied between t_{n-1} and t_n .” If the real time T that the wave needs to travel a distance of Δx is not an integer multiple of the simulation time step Δt , the observed effect will occur.

Next, it is worth looking at the impact of the simulation window dimensions on the wave front velocity. Up to now, I have always used $x_{\min} = -100$ and $x_{\max} = 50$ as initial boundaries for the lattice. In Fig. 3.6 the measured velocity is plotted as obtained by larger simulation areas $x \in [-100, 100]$ and $x \in [-250, 250]$. Besides the fact that both single-site and tanh i.c. now converge upwards, another effect can be observed: the velocity seems to converge towards larger values in larger simulation areas (again, note the different time scales).

This is further examined in a comparison of several different simulation area setups using a slightly modified measurement process. Instead of just doing one measurement process after the equilibration, both lasting as long as the simulation window has shifted the distance of ten times the simulation area length, several consecutive measurements were performed, each of which lasted as long as the window needs for moving five times the area length. The results (Fig. 3.7, left) show that the wave front velocity depends on the simulation area in front of the wave but not on the area size behind the wave front, which is consistent with the theoretical consideration at the end of Sec. 2.5.1. The first non-unity density of infected people (i.e., the first site with density less than 0.99999999) was always found at position $x = -28.6$ behind the wave. Interestingly, though also consistent with the theory, even reducing the simulation area down to $x_{\min} = -20$, i.e., cutting off parts of the wave, does not impact the measured velocity.

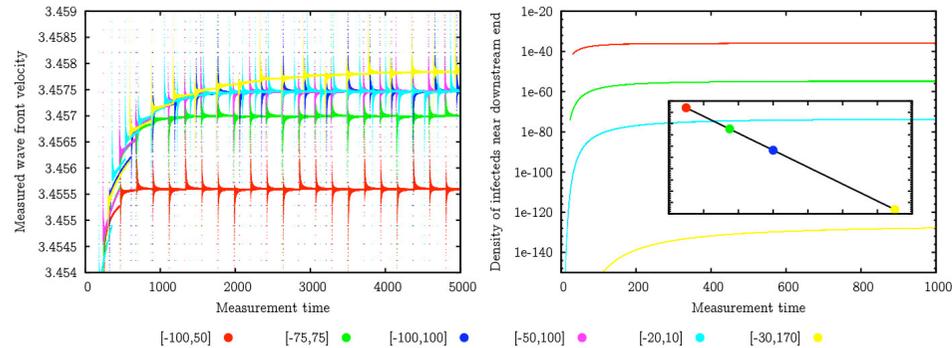


Figure 3.7: Multiple consecutive measurements of the wave front velocity in different system sizes. The inset in the right figure shows the logarithm of the equilibrated values plotted against the size of the simulation area ahead of the wave, i.e., $j(x_{\max} - 1, t = 5000)$.

The right plot of Fig. 3.7 shows the density of infecteds at a site near the downstream end of the simulation window (to be specific, it is the site with $x = x_{\max} - 1$ where x_{\max} of course changes with each window shift). As visible also from the left plot, the equilibration process takes longer for larger x_{\max} . The inset, showing the equilibrated density for the different lattice sizes, reveals that for large ξ the density $j(\xi)$ falls exponentially. Thus, the equilibrated wave, in theory, extends over the whole x -axis. However, in the finite simulation area, the boundary conditions impose that $j(x > x_{\max}, t) = 0$ and therefore significantly lower the (theoretical) mass of infecteds ahead of the wave, which, following the argumentation for the increasing wave speed with larger Δx , leads to a slow down of the wave front.

Applying the insights gained from the previous simulations, Fig. 3.3 has been reproduced using a larger simulation area and longer equilibration times, exhibiting significant accuracy improvements as shown in Fig. 3.8. Furthermore, the same simulations have been performed using the Runge-Kutta method (Fig. 3.9). As expected, the velocity converges much faster as $\Delta t \rightarrow 0$. Unfortunately, however, the Runge-Kutta method only improves accuracy to the time discretization, while the spatial discretization is untouched and exhibits the same convergence behavior as with the Euler method—again, not a big surprise. Also note that the Runge-Kutta method does not influence the stability criterion. Thus, for a small Δx that provides good accuracy, still a small Δt has to be chosen to ensure numerical stability and the Runge-Kutta method cannot unfold its full power.

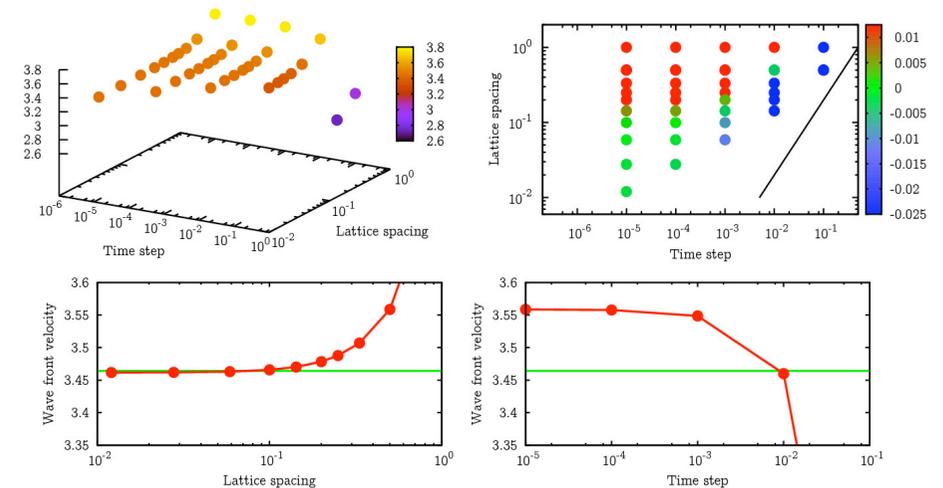


Figure 3.8: Accuracy of the speed measurement: Euler, revised. As in Fig. 3.3, but now with simulation area $x \in [-30, 100]$, equilibration for $10N$ window shifts and measurement for $5N$ window shifts.

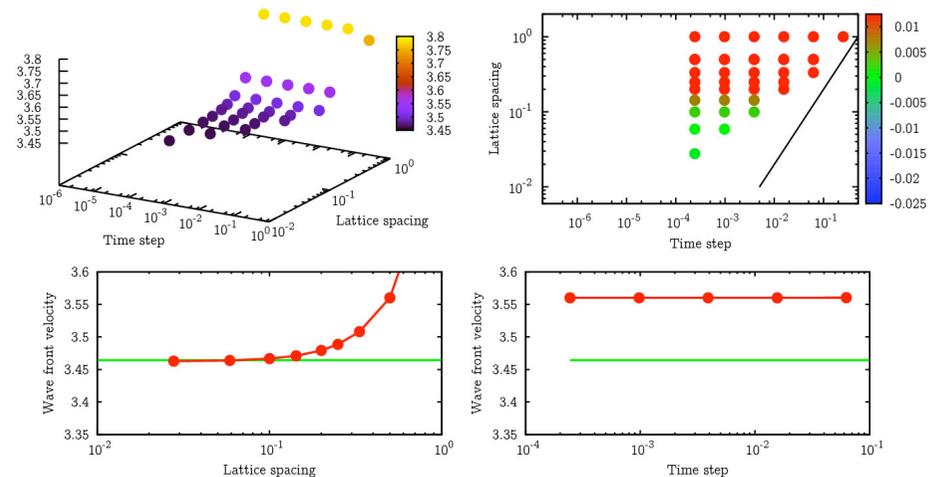


Figure 3.9: Accuracy of the speed measurement: Runge-Kutta. The slices in the bottom row are for $\Delta t = 2.4 \cdot 10^{-4}$ (left) and $\Delta x = 0.5$ (right).

4 Results

This page intentionally left blank.¹

¹Three design principles are the foundation of this chapter: First, since this chapter should be free of generic methodological details and interpreting statements, and my results can be nicely presented graphically, there is no need for any accompanying text. Figure captions describe what is shown in the plots and give specific methodological details where necessary. Second, regarding the question of whether to show all results obtained in the study or only the most important ones, I decided to provide a comprehensive survey while avoiding excessive visual repetition by omitting qualitatively similar results from the plots. Third, I chose a compact presentation form that visually groups strongly related results. Especially the double page 36/37 might appear overcrowded at first sight, but I am confident that, after reading the figure captions and understanding what the plots display, the reader will appreciate the compact and comprehensive overview this highly structured presentation provides.

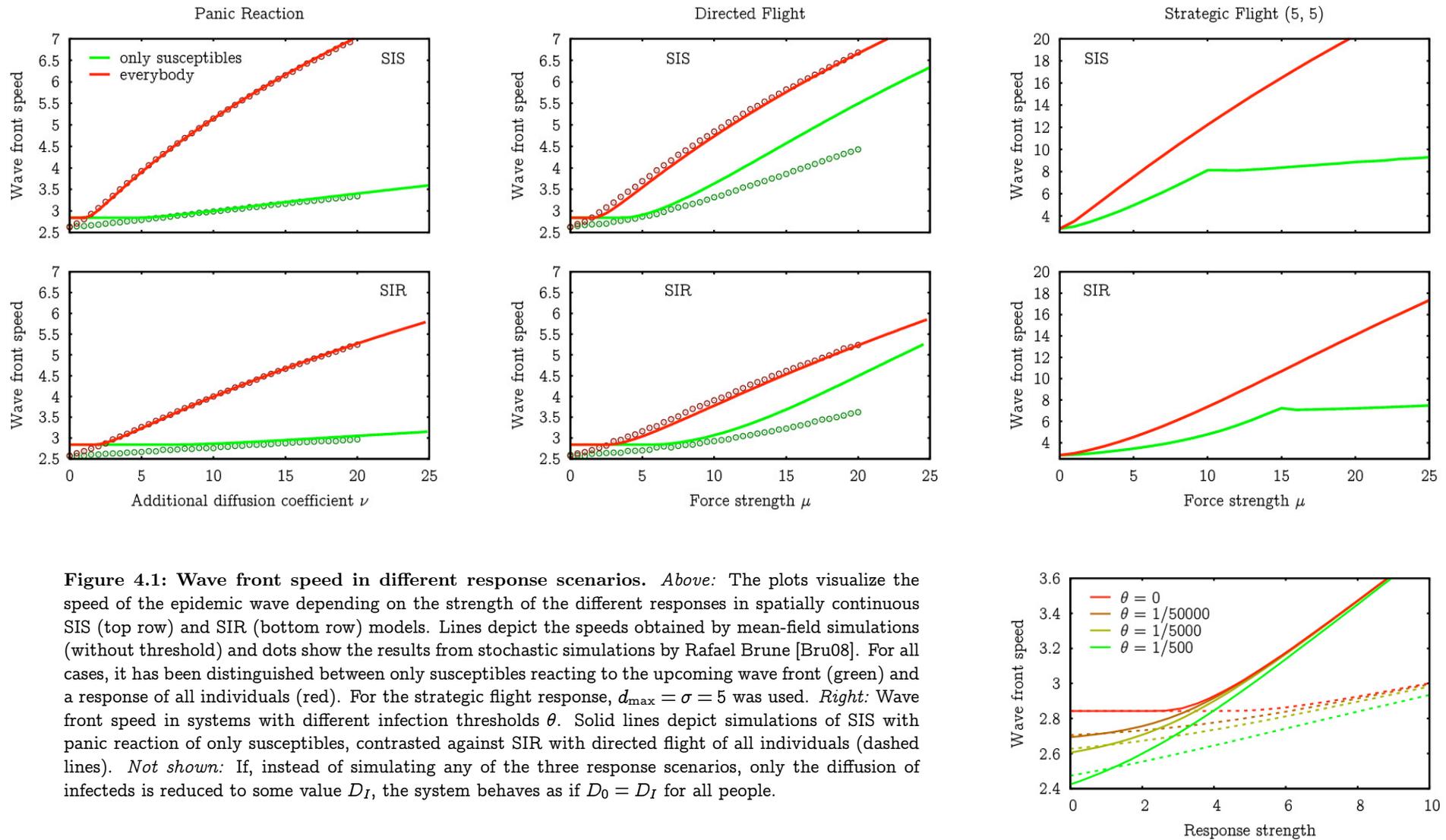


Figure 4.1: Wave front speed in different response scenarios. *Above:* The plots visualize the speed of the epidemic wave depending on the strength of the different responses in spatially continuous SIS (top row) and SIR (bottom row) models. Lines depict the speeds obtained by mean-field simulations (without threshold) and dots show the results from stochastic simulations by Rafael Brune [Bru08]. For all cases, it has been distinguished between only susceptibles reacting to the upcoming wave front (green) and a response of all individuals (red). For the strategic flight response, $d_{\max} = \sigma = 5$ was used. *Right:* Wave front speed in systems with different infection thresholds θ . Solid lines depict simulations of SIS with panic reaction of only susceptibles, contrasted against SIR with directed flight of all individuals (dashed lines). *Not shown:* If, instead of simulating any of the three response scenarios, only the diffusion of infecteds is reduced to some value D_I , the system behaves as if $D_0 = D_I$ for all people.

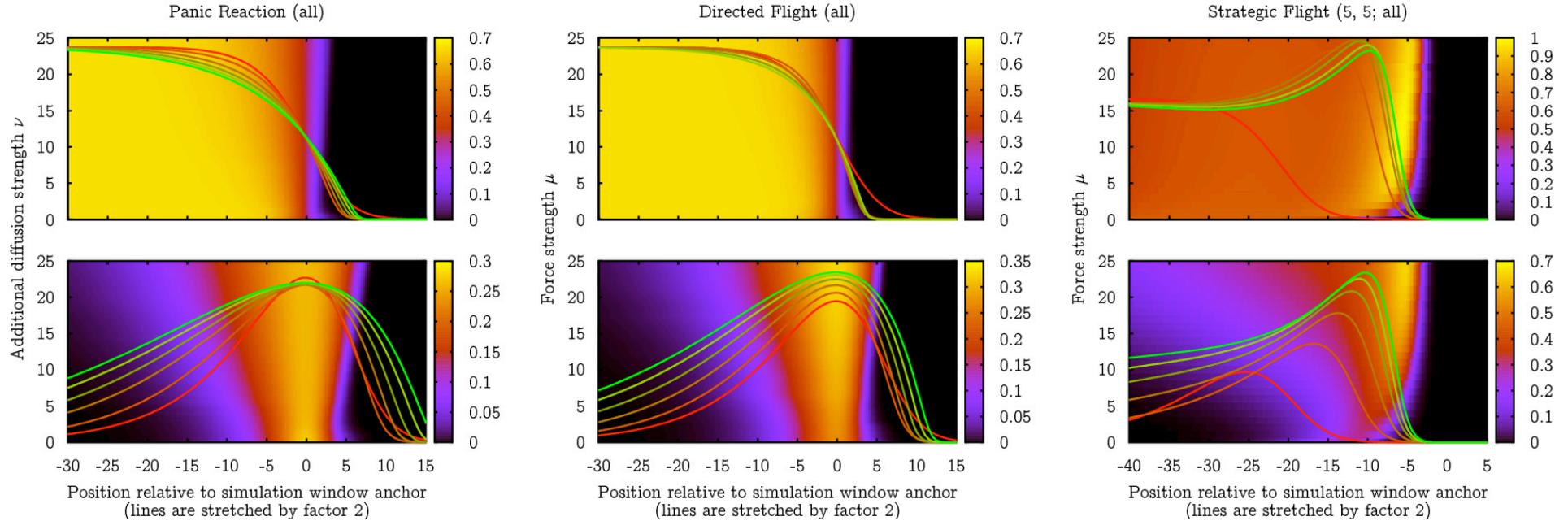
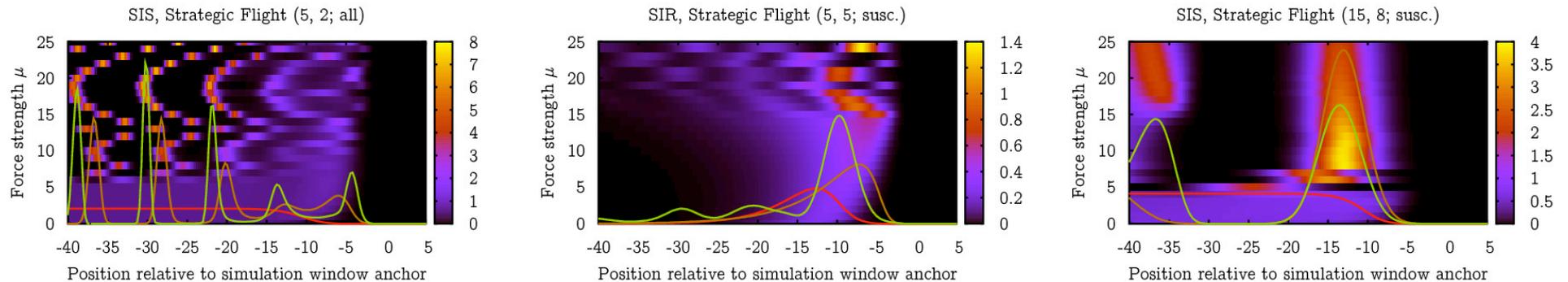
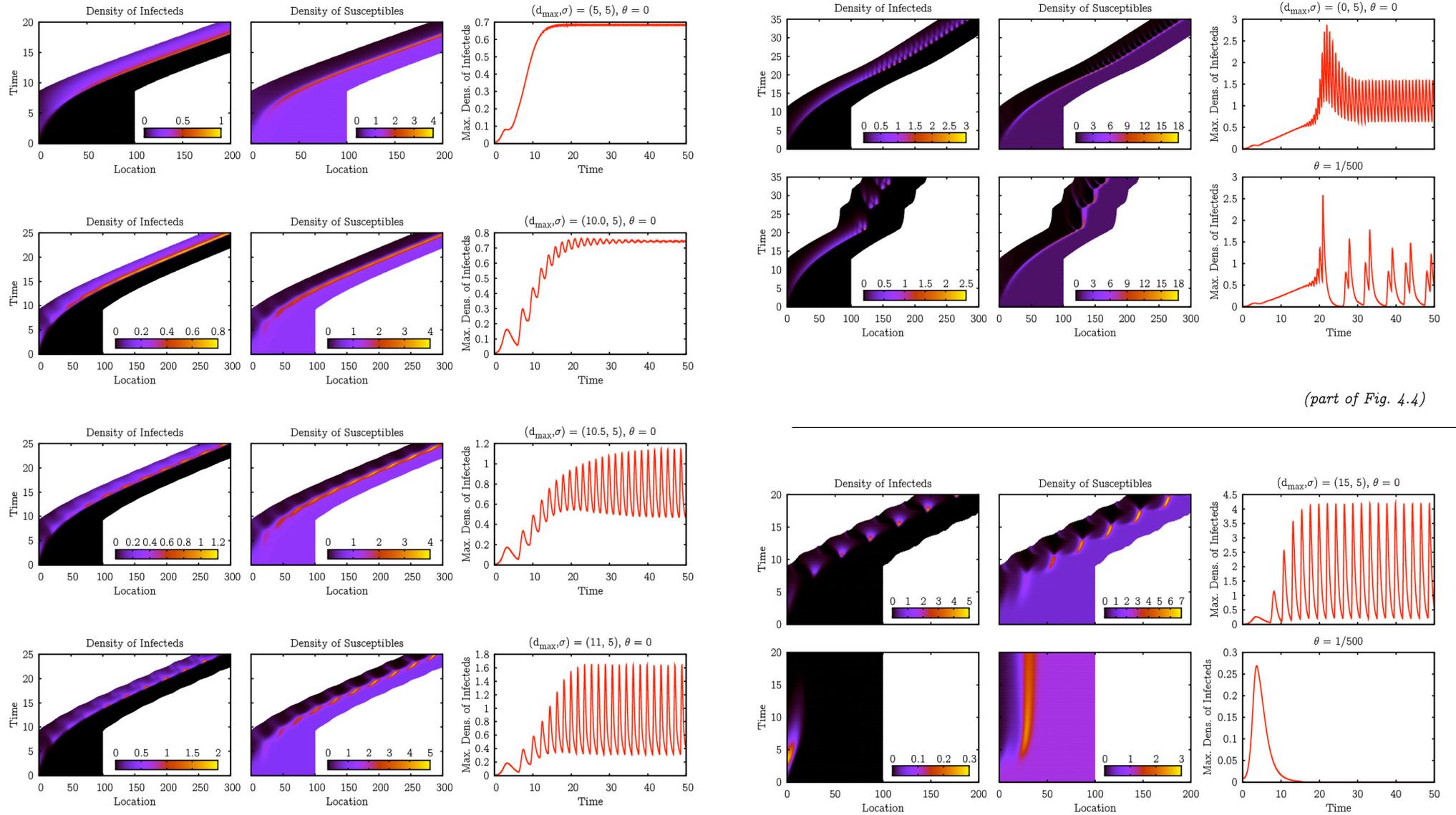


Figure 4.2: Wave front shape in different response scenarios. *Above:* The plots visualize the equilibrium form of the epidemic wave depending on the strength of the different population responses of all individuals in spatially continuous SIS (top row) and SIR (bottom row) models. Colors code the density of infected people. On top of the color maps, the wave front shape has been plotted for different values of the response strength parameter from 0 (red line) to 25 (green line). For the lines, the y -axis (density of infecteds) is given by the color band labels and the x -axis is stretched by a factor of 2, i.e., the lines are plotted for $x \in [-15, 7.5]$. *Below:* Same as above, but for different pa-

rameters (d_{\max}, σ) of the strategic flight response and also with only susceptibles reacting (center and right). Contrary to above, here the lines are plotted unstretched for $x \in [-40, 5]$. Also, different to panic reaction and directed flight, in simulations with the strategic flight the simulation window moved everytime the density of infecteds at the site with (initially) $x = 0$ exceeded 10^{-5} . Furthermore, the plots show only snapshots of the wave front shape $j(\xi)$, which is not necessarily time-independent in the strategic flight scenario. *Not shown:* With panic reaction or directed flight by susceptibles only, the effects are qualitatively the same as above.





(part of Fig. 4.4)

Figure 4.3: Initial SIR outbreak dynamics in the strategic flight scenario with response of all individuals. The plots show the spatio-temporal evolution of the density of infecteds $j(x_i, t)$, density of susceptibles $s(x_i, t)$, and maximum density of infecteds $\hat{j}(t) = \max_i j(x_i, t)$ of SIR dynamics with $\alpha = 3$ and $\beta = 1$ and strategic flight response of all individuals with strength $\mu = 25$. The left column shows the unthresholded dynamics for $\sigma = 5$ and different values $d_{\max} = 5, 10, 10.5, 11$. The corresponding thresholded systems with $\theta = 1/500$ produce similar results. In the right, $(d_{\max}, \sigma) = (15, 5)$ is shown for both the unthresholded and thresholded case.

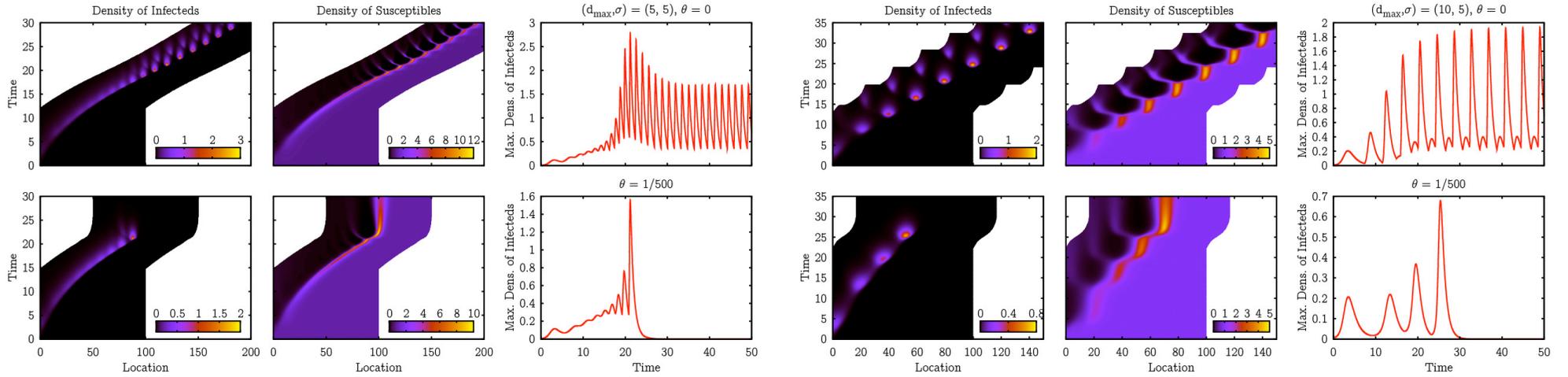


Figure 4.4: Initial SIR outbreak dynamics in the strategic flight scenario with response of only susceptibles. As in Fig. 4.3 on the facing page, spatio-temporal dynamics of the SIR model are visualized. Both thresholded and unthresholded systems are shown for $\sigma = 5$ and $d_{\max} = 5, 10, 15$ (from left to right, $d_{\max} = 5$ being found on the facing page).

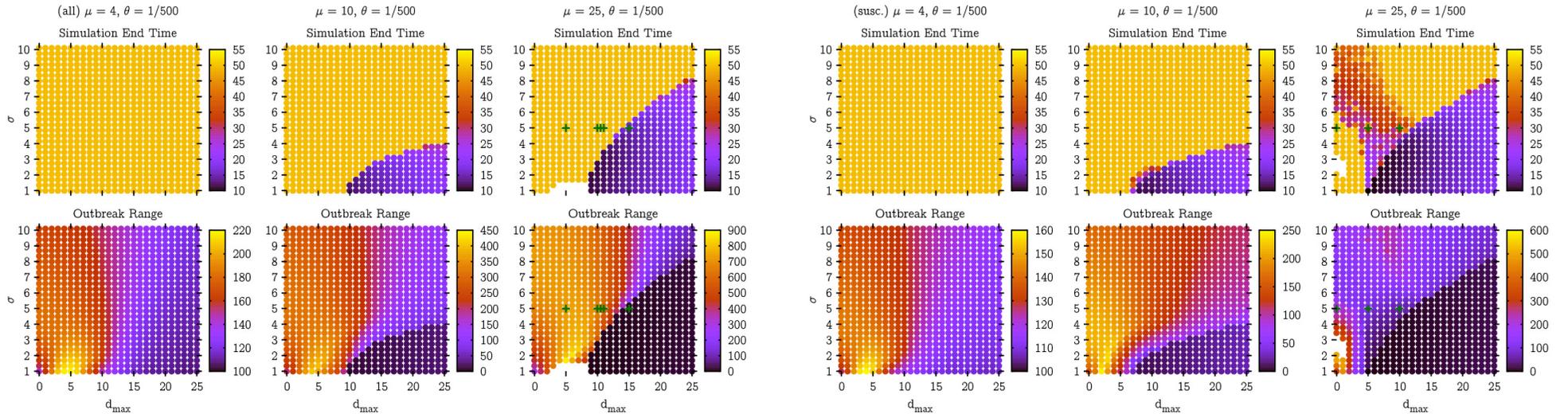
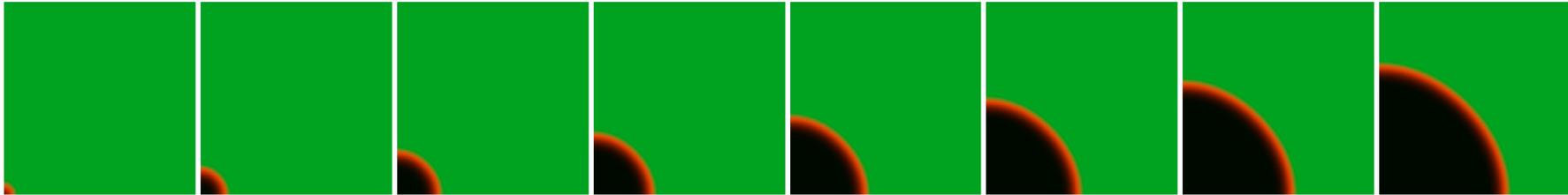
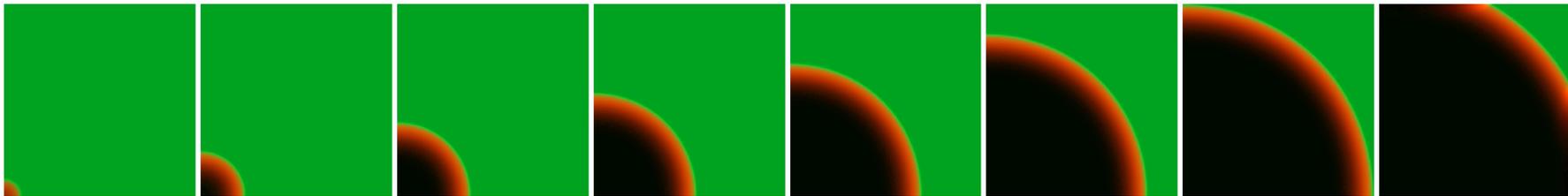


Figure 4.5: Parameter scan of the strategic flight scenario. For different response strengths $\mu = 4, 10, 25$ and $(d_{\max}, \sigma) \in [0, 25] \times [0, 10]$ simulations have been carried out in a simulation area $x > 0$ using the SIR model with strategic flight response with threshold $\theta = 1/500$ until either $t > 50$ or the disease was extinct ($j(x_i, t) < \theta$ for all sites i). The plots visualize the simulation end time and the outbreak range ($\max\{x_i | j(x_i, t) \geq \theta \text{ for any } t\}$) for response of all individuals (left half) or response of only susceptibles (right half). The green crosses mark the parameter sets whose dynamics are shown in detail in Figs. 4.3 and 4.4. Missing dots (for $\mu = 25$) are due to numerical instabilities.

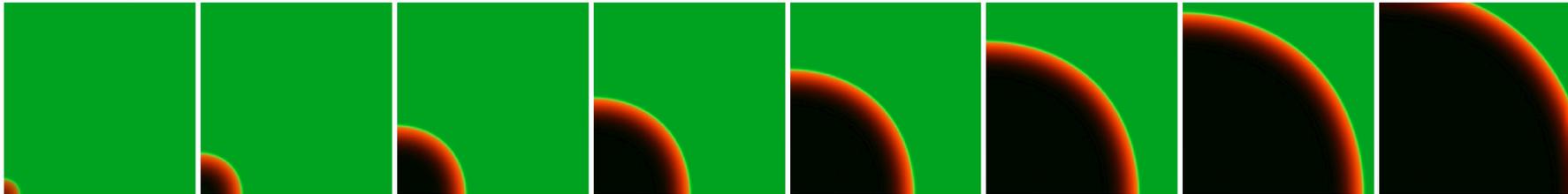
No Response



Panic Reaction



Directed Flight



Strategic Flight

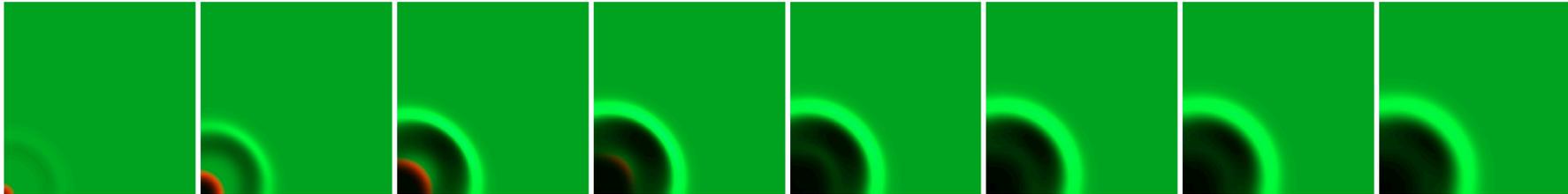


Figure 4.6: Qualitative picture of the outbreak dynamics in different response scenarios. Each horizontal image series present snapshots of the density of susceptibles (green) and infecteds (red) for different times (early pictures left, later pictures right). The simulations were carried out on a two-dimensional lattice using the SIR model with $\alpha = 3$ and $\beta = 1$. The scenario parameters were chosen large enough to distinctively show their effects. In each case, a small amount of infecteds has been placed in the lower left corner as the initial condition. Note that the density of recovered is not visualized, thus a black spot in the image may indicate a total population density of 1 but where almost all people are recovered (as inside the circular epidemic wave) or a vanishing total population density (as in the ring ahead of the epidemic wave in the strategic flight scenario).

5 Discussion

The results presented in the previous chapter point to some general conclusions, which I will state in the following. Also, I will give explanations to understand the model's mechanisms producing the observed effects. Afterwards, the relevance of the presented work is briefly discussed and the chapter closes with a short outlook to future work on this topic.

5.1 Conclusions

Regarding the equilibrium dynamics (Figs. 4.1 and 4.2), we can clearly see that the panic reaction and directed flight responses consistently lead to an increase of the wave front velocity. There are two mechanisms causing the acceleration. First, infecteds running away from an infected location lead to a faster physical transport of infecteds and thus, obviously, to a faster disease spread. Second and more subtle, since the response effect is local (either locally increased diffusion or a drift along the wave front that vanishes at the front's onset), a crowd of susceptibles will gather close to the wave front. Due to the increased density, the infection dynamics will run faster in this area and therefore new infected people, which can then diffuse further downstream, are available earlier than in the scenario without any population response. The second effect is of course the only acceleration mechanism in the case of only susceptibles reacting to the upcoming wave front.¹

¹The observing reader may have noticed that for the panic reaction of all individuals my results perfectly agree with the stochastic simulations by Rafael Brune (except for small μ , which can be explained by an implicit infection threshold θ present in Rafael's system). However, for the directed flight of all individuals, the results slightly differ, which can be explained by subtle differences in the implementation details (Rafael used different discretizations of the first derivative to calculate the gradient force, depending on the drift direction (as shown in Sec. 2.6.4), while in my simulations the centered difference was used in all cases). More severely, however, is the difference of the results in the cases where only susceptibles are reacting—particularly visible in the directed flight scenario but also present in the panic reaction scenario, though, due to the lower wave front acceleration, not as distinct—for which I cannot provide an explanation (yet). It is possible—though rather unlikely, of course—that either I or Rafael made a mistake in our implementations, or maybe a true stochastic effect that we have not yet understood appears in this scenario.

Also, we can see from the results that an infection threshold has significant impact on the wave front velocity for small response strengths. This can be explained by the fact that in thresholded systems diffusion has to carry a defined mass of infecteds to a currently susceptible site before the exponential growth due to infections can take over—contrary to systems without threshold where already an infinitesimally small amount of infecteds is exponentially amplified. For larger response strengths, the larger front velocity effectively leads to faster diffusion and decreases the time between the first infecteds entering a site and the density at that site exceeding the infection threshold. Thus, the wave front velocity in a thresholded system converges to the velocity of the unthresholded system for large response strengths (cf. Fig. 4.1, bottom right).

The snapshots of the equilibrium wave front shapes (Fig. 4.2) show that the two local response scenarios lead to a distortion of the wave front, particularly a steeper onset in the SIS model, caused by the faster diffusion and infection dynamics as explained above, and larger wave form content $\int j(\xi) d\xi$ in the SIR model. Thus, more people are affected by an SIR-like disease at the same time. Obviously, due to the larger front velocity, the total number of affected people in either SIS or SIR model at a given fixed time t , $\int j(x, t) + r(x, t) dx$, is consistently increased by the population's response.

The same conclusions apply to the strategic flight scenario for certain parameter sets (d_{\max}, σ), but the bottom part of Fig. 4.2 indicates that this scenario exhibits a richer dynamics than the local responses, as discussed in the following. In particular, for some parameter sets, the wave form does not converge to a stable solution $j(\xi)$ but exhibit spatio-temporal oscillations.² Also, in the initial

²In other words, if the function $j(\xi)$ is interpreted as a dynamical variable $\mathfrak{J}(t) = j(x - ct, t)$ with appropriate constant c , in the corresponding phase space, which contains all functions parameterized by, e.g., Taylor or Fourier series coefficients, there exists a function \mathfrak{J}^* for the panic reaction and directed flight scenario such that $\lim_{t \rightarrow \infty} \mathfrak{J}(t) = \mathfrak{J}^*$, while for the strategic flight scenario it seems that there is a closed set of functions $\mathfrak{J}^\# = \{\mathfrak{J}^*(\tau)\}$, which can be parameterized by a single variable, such that the smallest open set containing $\mathfrak{J}(t)$ and $\mathfrak{J}^\#$ converges to $\mathfrak{J}^\#$ for $t \rightarrow \infty$. Obviously, since stable solutions in the strategic flight scenario exist for certain parameter sets, $\#\mathfrak{J}^\# = 1$ is possible.

outbreak (non-equilibrium) dynamics, the strategic flight scenario exhibits a large behavioral variety, strongly depending on the chosen parameter set (d_{\max}, σ) and the infection threshold θ . As visible in Fig. 4.5, with $\theta > 0$ the parameter space splits into two phases at some critical response strength μ_c . One phase exhibits traveling waves and one extinction of the disease. How can this happen? If the infection wave runs into an evacuated region with a low density of susceptibles, the infection dynamics do not exhibit exponential growth of infected people, but exponential decay.³ Thus, in the evacuated region a disease is suppressed, i.e., it is locally extinct. However, diffusion will always carry ridiculously small amounts of infecteds beyond the evacuated area and thus triggers a new wave front after the bottleneck, since there infecteds are exponentially amplified again. This can be prevented by setting a non-zero infection threshold that requires a defined amount of infecteds to trigger exponential growth. If the decaying wave front upstream of the bottleneck cannot provide enough infecteds to quickly transport the required amount beyond the bottleneck, the disease will globally die out.

The transition between both phases can be understood by looking at the non-thresholded dynamics of the system. The space-time plots in Fig. 4.3 show, that for fixed σ and increasing d_{\max} an oscillation appears in the temporal evolution of the wave front whose amplitude is growing with d_{\max} . Since the oscillations are conveyed to the density of susceptibles via the population's response, for sufficiently large oscillation amplitudes the total population density ahead of the wave will drop below ρ_c . Thus, with a non-zero threshold, the wave propagation can be stopped.

If not all population's individuals are reacting to the upcoming wave front but only the susceptibles (Fig. 4.4), the dynamics are further enriched. In addition to the two phases observed in the scenario with everybody responding, a third phase, in which growing oscillations can lead to a disease extinction, develops for large μ .⁴ In contrast to the extinction phase also observed in the scenario with everybody responding, where a wave hits a single impassable bottleneck, in this new phase the wave passes the bottleneck and strikes into a crowd of susceptibles beyond the evacuated area. Since this crowd exhibits a larger population density than upstream of the bottleneck, the density of infecteds will grow to larger values than before, leading to a more intense response and a deeper bottleneck downstream. If this bottleneck is also passed, the next bottleneck will be even deeper and so

on, until, finally, an impassable bottleneck is created. Of course, in this phase, a yellow spot in Fig. 4.5, indicating that the simulation was aborted at $t = 50$ with infecteds present in the system, does not imply that a stable travelling wave will emerge.

5.2 Relevance

This study is telling the reader that if you have a system like one of those introduced in Chapter 2, it will behave in a way such that, if investigated with the methods presented in Chapter 3, the effects shown in Chapter 4 can be observed, which might be understood by the arguments presented in Section 5.1. The study does not tell you whether your system of interest actually is of this type. To decide this, instead of investigating *intrinsic* properties of the purely mathematical model, its basic axioms (assumptions) have to be empirically checked for consistency with what most people call “reality.”

Clearly, the models presented in this work fail this check at least with respect to what they claim to represent. As already mentioned in Chapters 1 and 2, the spatially continuous diffusion framework does not correctly capture modern human travel behavior and therefore cannot correctly predict the global dynamics of a spreading disease, since it has been shown (e.g. [Huf04]) that modern travel means have significant impact on this. However, to be precise, it cannot correctly predict the effects of a population's response in the real world, if the modern travel opportunities have any relevance to the overall travel behavior of the population of interest. An argument can be made that the simple diffusion framework might be valid, i.e., consistent with reality, in the case of, e.g., a bio-terroristic attack using a highly infectious disease. Here, the time-scale of the disease spread might be faster than the time-scale of the long-range travel behavior—or the long-range travel means such as air traffic and intercity trains might have been shut down by the authorities. Also, simple diffusion models are applicable to many problems in ecology [Bul02], where they, for example, describe the migration behavior of animals.⁵ However, it is to be determined if the formulated response scenarios can be found in ecological systems—or if they correctly represent human behavior in case of an outbreaking disease.

In conclusion, the models studied in this thesis serve more as a toy model than being useful as a realistic simulation tool. They can be used to investigate the

⁵At least of those that have neither been gifted with fast long-range travel means nor engineered ones themselves.

³Remember from Sec. 2.3 that there is a critical density $\rho_c = \beta/\alpha$ below which the trivial fixed point $j^* = 0$ becomes the only, stable fixed point.

⁴The simulations have been repeated with smaller discretization parameters Δt and Δx for a selected parameter regime and produced similar results. I am therefore confident that the oscillations are not a numerical artifact.

effects of certain phenomenologically motivated population response scenarios in the simplest framework of spatially extended reaction-diffusion systems.

The high-spirited reader might conclude from this study, that—given a real world scenario that is correctly described by simple diffusion—people should stop running away if the disease is coming too near, since this response behavior, corresponding to the strategic flight scenario, will prevent a large-scale outbreak. I want to emphasize that this is not the case—and cannot be by principal reasons. This study, as any scientific study, tells the reader that certain systems *do* behave in certain ways and not that certain systems *should* behave in a certain way. However, it might be interesting to transgress the scientific boundary and discuss the ethics of forcing people to a strategic flight response.

The strategic flight implies that people run away from the dangerous disease only if they are sufficiently far away. They have to stop running away and surrender themselves to the risk of infection if the wave front is too near. Obviously, since the whole point of this behavior is to stop the outbreak and prevent more people becoming infected, these people “sacrifice” themselves to save the larger population. If people are forced to do this, their lives—or at least their health—is traded for the lives or health of others, rendering the “sacrificed” persons a means to an end. This contradicts Kant’s definition of human dignity and is thus unacceptable. Two ways might provide an escape route from this dilemma: First, instead of forcing people into a strategic flight, they could be informed about this behavior and encouraged to do so voluntarily. In my opinion, however, it is difficult to cut the line between “forcing” and “encouraging”—if there is any relevant difference at all. The second option is the possibility of self-protection of “sacrificed” people. If people can protect themselves against infection by, e.g., strictly staying inside their homes or wearing protective masks, the information and encouragement approach might appear less wrong. However, the argument can only be made if reliable self-protection is available—forcing or encouraging people to stay if they only have, say, a 90% chance of not being infected, is probably considered careless as well. However, assuming that reliable self-protection is available, one can ask why a strategic flight response is necessary at all in this case: Why should anybody want to run away anyway if they can easily protect themselves at home? Probably the only case in which an enforced strategic flight is useful and ethical is when reliable self-protection is available in the threatened area but not in sufficient quantity to protect everybody. In this case, people further away can be forced to move away from the outbreak region while the limited self-protection devices/facilities are moved towards the outbreak region to protect the people that have not fled.

5.3 Outlook

There are some interesting possible follow-ups to this study. The most important is to replace the simple diffusion with a more realistic model of human travel behavior and see if the response scenarios still exhibit the same or at least similar effects. Two general approaches can be undertaken to achieve this.

The first option is probably the easiest and most straight-forward. The panic reaction scenario is canonically defined in a network and the directed flight scenario can also easily be generalized to arbitrary networks by choosing a proper discretization scheme for the gradient force, as shown in Sec. 2.6.4. However, a meaningful network equivalent of the strategic flight has still to be thought of. Furthermore, it is hard to measure a wave front velocity in an arbitrary network. Thus, new measures of response impact have to be defined. Interesting and relevant quantities might be the total number of people affected by the disease, the number of infected people at a given time, and their spatial extent at a given time. They indicate the overall “seriousness” of the pandemic, the momentary load on medical facilities, and the range over which possible physical counter-measures like vaccines have to be distributed. For measuring the last quantity it is of course necessary to have a geographically embedded network. Empirical data for estimating human travel rates between cities or countries can be obtained by various means such as evaluating air traffic data [Huf04] or databases of, for example, the dollar bill tracking game “Where’s George?”⁶ [Bro06] or the real-life adventure game “Geocaching”⁷.

The second option is a somewhat different ansatz. The regular diffusion is replaced with superdiffusion, i.e., agents on a linear chain or regular grid do not perform random jumps of equal step size, but the length of each jump is randomly distributed as well. In particular, the jump lengths might be distributed according to a power law. Then, the resulting trajectories of superdiffusive random walkers exhibit a lot of very small steps and a few very large jumps. These superdiffusive long-range jumps can be interpreted as train rides or flights. The advantage of this approach is that it can also be defined in the spatially continuous framework, making it unnecessary to reformulate the response scenarios.

Furthermore, in both approaches, a fourth response scenario, in which the probability of a person performing a long-range jump is influenced by the disease prevalence at the person’s current location or in its environment and at the person’s potential jump target or in its environment, can be investigated.

⁶<http://www.wheresgeorge.com/>

⁷<http://www.geocaching.com/>

Bibliography

- [Bro06] Dirk Brockmann, Lars Hufnagel, and Theo Geisel: *The Scaling Laws of Human Travel*. *Nature* **439**(7075):462–465 (2006). 41
- [Bru08] Rafael Brune: *A Stochastic Model for Panic Behaviour in Disease Dynamics*. Master’s thesis, Georg August University Göttingen (2008). 21, 23, 34
- [Bul02] James M. Bullock, Robert E. Kenward, and Rosie S. Hails: *Dispersal Ecology*. Cambridge University Press (2002). 40
- [Fey74] Richard P. Feynman: *Cargo Cult Science*. *Engineering and Science* **37**:7:10–13 (1974). 19
- [Gar85] Crispin W. Gardiner: *Handbook of Stochastic Methods*. Springer New York (1985). 1, 9
- [Gil76] Daniel T. Gillespie: *A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions*. *Journal of Computational Physics* **22**:403–434 (1976). 17
- [Heu00] Karl Heun: *Neue Methode zur approximativen Integration der Differentialgleichungen einer unabhängigen Variablen*. *Zeitschrift für Mathematik und Physik* **45**:23–38 (1900). 24
- [Huf04] Lars Hufnagel, Dirk Brockmann, and Theo Geisel: *Forecast and Control of Epidemics in a Globalized World*. *Proceedings of the National Academy of Sciences* **101**:15124 (2004). 40, 41
- [Ker27] William Ogilvy Kermack and Anderson Gray McKendrick: *A Contribution to the Mathematical Theory of Epidemics*. *Proceedings of the Royal Society London A* **115**:700–721 (1927). 1, 7
- [Kut01] Martin Wilhelm Kutta: *Beitrag zur näherungsweise Integration totaler Differentialgleichungen*. *Zeitschrift für Mathematik und Physik* **46**:435–453 (1901). 24
- [Mur89] James D. Murray: *Mathematical Biology*. Springer Berlin Heidelberg (1989). 1, 19
- [Pre02] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery: *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press (2002). 1
- [Run95] Carl Runge: *Über die numerische Auflösung von Differentialgleichungen*. *Mathematische Annalen* **46**(2):167–178 (1895). 24
- [Str94] Steven H. Strogatz: *Nonlinear Dynamics and Chaos*. Westview Press (1994). 19

Acknowledgements

Bernd Blasius (ICBM Oldenburg)

for posing the initial question that lead to this study

Dirk Brockmann

for his excellent supervision and new perspectives on science

Rafael Brune

for a great collaboration on this topic

Theo Geisel

for accepting me as his graduand and for providing a great work environment that leaves nothing to be desired except a constantly changing beautiful sweeping landscape outside the window

metronom Eisenbahngesellschaft

for filling this gap

Vitaly Belik, Vincent David, Alejandro Morales Gallardo, and Benjamin Schwenker

for useful, interesting and entertaining discussions

Arne Kesting and Martin Treiber (TU Dresden)

for a great time with traffic data and a solid introduction to scientific work

Christin Kalkert, Katharina Wilhelm, and Manuela Wengelnik

for reviewing the manuscript and providing valuable comments

My Parents