## PHD DISSERTATION

Investigating space-dependentepidemic models andthe collapse of Easter Island
Takács Bálint Máté
Supervisors:
Faragó István
Doctor of the Hungarian Academy of Sciences
ELTE Department of Applied Analysis and Computational MathematicsandHorváth RóbertAssociate professorBME Department of Analysis
Eötvös Loránd University, Faculty of Science Doctoral School in Mathematics Director: Jordán Tibor
Doctor of the Hungarian Academy of Sciences
Doctoral Program in Applied Mathematics Director:
Karátson János
Doctor of the Hungarian Academy of Sciences ..... DOI: 10.15476/ELTE.2021.186

## Contents

I Epidemic models ..... 4
$1 \quad$ Space dependent epidemic models ..... 7
1.1 Existence of the analytic solution ..... 9
1.2 Properties of the analytic solution ..... 12
1.3 Numerical approximation of the integral solution ..... 16
1.4 Spatial discretization ..... 18
1.4.1 The semi-discretized system ..... 19
1.5 Discretization in time: the method of Euler ..... 25
1.6 Discretization in time: Runge-Kutta methods ..... 27
1.6.1 Strong stability preserving Runge-Kutta methods ..... 28
1.6.2 The application of SSP-RK methods ..... 31
1.6.3 Numerical results ..... 32
1.7 Discretization in time: Operator splitting techniques ..... 36
1.7.1 Technical tools ..... 39
1.7.2 Sequential splitting ..... 41
1.7.3 Weighted sequential splitting ..... 47
1.7.4 Strang splitting ..... 52
1.7.5 Numerical results ..... 54
1.8 Conclusions ..... 60
2 Epidemic models with latency period ..... 62
2.1 Existence and properties of the analytic solution ..... 63
2.1.1 Qualitative properties of the analytic solution ..... 65
2.2 Numerical approximation ..... 66
2.2.1 The spatially discretized models and their properties ..... 66
2.2.2 Time discretization: explicit Euler method ..... 68
2.2.3 Time discretization: Application of Runge-Kutta methods ..... 70
2.3 Numerical experiments ..... 71
2.3.1 Construction of the test problem ..... 71
2.3.2 Sharpness of the time step bounds ..... 73
2.3.3 Comparison of the cases with different delay parameters ..... 75
2.4 Conclusions ..... 75
3 Possible further extensions ..... 77
II Modeling the collapse of Easter Island ..... 79
4 One dimensional model ..... 81
4.1 Discretized model ..... 81
4.1.1 The effect of the diffusion of trees ..... 85
4.1.2 The choice of the parameters $f$ and $N$ ..... 90
4.1.3 Numerical solution of the equations ..... 94
4.1.4 Global stability of the equilibrium points ..... 99
4.2 Continuous model ..... 106
4.2.1 Construction of the model ..... 107
4.2.2 The effect of the diffusion of the trees ..... 108
4.3 Conclusions ..... 112
5 Two-dimensional model ..... 114
5.1 Construction of the model ..... 115
5.2 The effect of the diffusion of trees ..... 117
5.3 Numerical solution ..... 120
5.4 Conclusions ..... 123
6 Possible further extensions ..... 125

## Rövid összefoglaló

Az utóbbi száz év talán leggyorsabban fejlődő matematikai ága a biomatematika, mely eredetileg populációdinamikai modellekből alakult ki. Jelen dolgozatban két ilyen modellcsaláddal foglalkozunk.

A dolgozat első része járványterjedési modellekkel foglalkozik - ezek közül az egyik leggyakrabban használt az ún. SIR modell. Ezen modell valamely tulajdonság (például megbetegedés) terjedését írja le egy adott populáción belül egy közönséges differenciálegyenletrendszer segítségével. Mivel a klasszikus SIR modell a térbeli terjedésről semmiféle információt se szolgáltat, ezért az eredeti modellt egy integrál hozzáadásával parciális integrodifferenciálegyenletté alakítjuk. Egy ilyen modell a disszertáció első fejezetében szerepel, míg a második fejezet ezt egy késleltetés beépítésével módosítja, majd a harmadik fejezet néhány további lehetséges módosítást tárgyal. Az említett szakaszok mindegyikében elsőként az egyenlet klasszikus megoldásának létezéséről és tulajdonságáról mondunk ki tételeket, majd olyan numerikus módszert konstruálunk, mely megőrzi ezen (a biológiai modellel szinkronban lévő) tulajdonságokat.

A dolgozat második részében a Húsvét-sziget (Rapa Nui) 16-17. századi ökológiai összeomlásának matematikai modelljét vizsgáljuk. A modell alapját azon történeti kutatások adják, melyek szerint nem csupán a felelőtlen emberi tevékenység, de a szigetre behurcolt patkányok is hozzájárulhattak a katasztrófához. A korábban ezen folyamatokat leíró modellt egy, a fák diffúzióját leíró tag hozzáadásával egészítjük ki. A módosítás oka, hogy a fák magjait a szigeten élő állatok kültakarójukon elszállíthatják, illetve a szél is mozgathatja ezeket. A dolgozat 4. és 5. fejezeteiben a fenti probléma különböző matematikai modelljeit (közönséges, egy térváltozós parciális és két térváltozós parciális differenciálegyenleteket) írjuk fel, és megvizsgáljuk a fák diffúziójának a sziget ökoszisztémájának stabilitására kifejtett hatását. A vizsgált modellek mindegyikében azt tapasztaltuk, hogy a patkányok jelenléte a szigeten előidézhette annak ökológiai összeomlását.

A fenti eredményeket hat megjelent, egy elfogadott és kettő benyújtott publikáció tartalmazza.

## Short summary

In the past century one of the fastest growing sub-fields of applied mathematics has been the area usually referred to as biomathematics, which originally developed from population dynamics. In this work we focus on two such sets of models.

The first part deals with epidemic models: one of the most used such model is the SIR model. This construction can describe the propagation of a given property (e.g. illness) among a set of individuals using an ordinary differential equation. Since the original model does not give us any information about the spatial distribution of the illness, we extend the original system using an integral term resulting in an integrodifferential equation. Such an extension is presented in Chapter 1, while in Chapter 2 it is further modified by adding a delay term to the equation, and in Chapter 3 some further extensions are mentioned. In the aforementioned sections we first state theorems about the existence and some properties of the solution, then construct such a numerical model which preserves these (biologically reasonable) properties.

In the second part we observe the ecological collapse of Easter Island (Rapa Nui) which took place during the 16 th and 17 th centuries. The ground for these models are the archaeological findings which propose that the cause of this catastrophe was not only the irresponsible consumption of trees by the inhabitants, but also the effect of the rats which were brought to the island by the settlers. We extend the previously proposed models by adding a term describing the diffusion of the trees: the reason for this is that the seeds could be transported on the fur of the animals or by the constant wind on the island. In Chapters 4 and 5 we examine several different mathematical models (ordinary, one space-variable partial and two space-variable partial equations) describing the aforementioned phenomenon, and observe the effect of the added tree diffusion on the stability of the island's ecosystem. In all of the models we find that the presence of the rats on the island could lead to the ecological catastrophe.

The aforementioned results are included in six published, one accepted and two submitted articles.

## Acknowledgment

First of all, I would like to express my deep gratitude to István Faragó and Róbert Horváth, my supervisors, for their continuous support guidance and enthusiastic encouragement during the four years of my PhD studies. I would also like to thank my other co-authors, namely Petra Csomós, Yiannis Hadjimichael, Illés Horváth and Dániel Keliger for the opportunity to work with them. Moreover, I would also like to mention all my colleagues at and outside of the department whose useful comments regarding my work helped to finish this work.

Finally, I wish to thank my family for their support and encouragement throughout my studies.

## Introduction

In the past century one of the fastest growing sub-fields of applied mathematics has been the area usually referred to as biomathematics. With the recent genomics revolution [7, 26, 39] there is a growing need for data analysts to be able to understand the data-rich information sets resulting from such processes 855. However, the topic that mathematical biology originally developed from was population dynamics. In this work we will focus on this latter one.

On the following pages we outline a short summary of the history of this branch of mathematics, focusing only on the most significant improvements. For further reading, we suggest the works of Bacaër [9], Daley and Gani [36] along with the references therein.

The very first instance in history (that the author knows of) when mathematics was applied to biological processes was in the book Liber Abaci written by Fibonacci in 1202 [95]. In his work the merchant from Pisa described a problem concerning the growth of the population of rabbits coming from a single pair, resulting in a difference equation and the famous sequence named after him. Unfortunately, most of his work was forgotten and the sequence was re-discovered later by Kepler and also by Daniel Bernoulli [18].

The 1662 book of John Graunt [56] can be considered the first work that observed an epidemic in a scientific way. Although it only calculated the percentage a citizen of London might die of a certain type of illness using death records from the previous decades, it is still the piece that started the mathematical description of pandemics. About 30 years later Halley [58] wrote a similar book about the deaths in the city of Breslau (now called Wrocław) in which (using the assumption that the population of the city remains constant) he could construct a life table estimating the number of people aged a given number. The first work that tried to estimate the number of ill and healthy individuals was written by Daniel Bernoulli [17]. In this he tried to find a solution to stop the smallpox pandemic which was a big problem among soldiers in the seven-year war. He argued that variolation (inoculation with live virus obtained directly from a patient with a mild case of the illness) would decrease the number of infectious people. Later his models were extended by several other authors (see e.g. [70]).

Although the model for the exponential growth of the human population is usually
associated with Malthus who popularised the concept in his 1798 book [76], the idea of using a geometric series had been already developed by Euler about 40 years before that [44]. While the former used his model mainly to demonstrate that the increase of living conditions cannot happen without the increase of food production, the latter used it to demonstrate the idea that mathematically it is possible to populate the Earth in only four millennia using a not too unrealistic growth rate of 6.25 percent (showing that the biblical story of Noah cannot be rejected on mathematical grounds). Another model was proposed by Verhulst [108] which had a solution tending to a constant value sometimes referred to as the capacity of the territory we examine the growth in. Verhulst called its model logistic, which was widely used in the following centuries: a similar idea is used in Chapters 4 and 5 of this work. A later extension of this model was done by Lotka nearly a century later [71, 91] in which he took into account the age dependence of fertility, resulting in an integral equation.

An important step further in the history of population dynamics was the discovery of the law of mass action. This theory, originally postulated in chemistry [19, 112, 113 ] said that for a homogeneous system, the rate of a chemical reaction is proportional to the active masses of the reacting substances. In population dynamics, the same idea can be applied in the following way: if the individuals in a population mix homogeneously, then the number of interactions between two observed subsets is proportional to the cardinality of each of these subsets. This simple idea then gave rise to a large number of new models and formed the basis for most of the differential equations which are used today to describe such phenomena. All of the following examples in this chapter use this idea in one way or another, along with the models described in the further parts of this work.

In the 1920s the predator-pray models proposed by Lotka [72, 73] and Volterra [110, 111 were the first to propose periodic solutions in a biological system, which gave rise to the in-depth observations of dynamical systems. Such methods are used in Part II of this dissertation.

The first model in the mathematical theory of epidemics which used differential equations was developed by Ross [87] during his work trying to stop malaria in India and other parts of the world. His main contribution to the theory was the idea that the disease is transmitted by mosquitoes, which was confirmed by experiments. Consequently, the reduction of the number of mosquitoes resulted in a significant decrease in the number
of deaths resulting from the illness. Later, Kermack and McKendrick [69, 79] developed their model which is still widely used today: Part【of this work also considers an extension of those ideas.

Although in this thesis we only consider deterministic models resulting in different types of differential equations (ordinary, partial, delay, integro-differential), it is worth mentioning that another powerful tool to model biological processes is stochastic methods. One of the largest influences on this area was the work of Yule [115] modeling the change of the number of species in time resulting from evolutionary processes. The derived model, sometimes called the "Yule process" is still widely used today not only to model evolution, but also to describe the growth of colonies of bacteria or to model the beginning of an epidemic [4]. Another incredibly influential stochastic model used to describe epidemics was introduced by Reed and Frost [1]. Their main idea was that infection is passed on by an adequate contact of an infectious and a susceptible person in a relatively short time. Each individual has a certain probability of infecting one of its contacts, making the model stochastic. The probability that an epidemic might emerge in this model is a product of a chain of binomials, hence it is sometimes called the chain binomial model. Since then, many have extended these ideas [12], resulting in sometimes discrete or continuous Markov chains [34, 35, 51]. Note though that stochastic modeling is usually suitable only if the size of the population is not too large; likewise, deterministic models are not favorable when we consider a small number of individuals. Since the observed systems in this work usually involve a large number of entities, we choose to use deterministic tools to describe them.

This dissertation consists of two parts: Part [ considers some epidemic models and proves that upon using a sufficient numerical model, the properties of these numerical solutions will be similar to the ones of the original continuous model. These results are included in the papers [33, 103, 104, 105. Part II deals with the problem of the collapse of the Easter island, and states some theorems about the stability of the corresponding system of differential equations, which were published in the articles [98, 99, 101, 102].

## Part I

## Epidemic models

During the millennia of the history of mankind, many epidemics have ravaged the human population. Since the plague of Athens in 430 BC described by historian Thucydides (one of the earliest description of such epidemics), researchers tried to model and describe the outbreak of illnesses. More recently, the outbreak of the COVID-19 pandemic revealed the importance of epidemic research and the development of models to describe the public health impact of major virus diseases.

One of the most frequently used tools in mathematics to model the spread of diseases is the SIR model. These models can be used to describe the spread of some feature among a group of individuals. Let us split our population into three categories: class $S$ contains those who do not have the property yet, class $I$ includes those who have the feature and they can also transmit it to others, and class $R$ has those who did have the property, but they do not have it any more. Note that these models can, and are used not only in epidemics, but also in several other fields: one such application is the modeling of forest fires, in which $S$ means the healthy trees, $I$ denotes the plants on fire and $R$ is the class of burnt down trees. For other applications, see e.g. [5], [20], [86], [109] - however, in this work we focus on epidemic models.

The aforementioned model can be written as the following first order system of ordi-
nary differential equations

$$
\left\{\begin{array}{l}
\frac{d S(t)}{d t}=-a S(t) I(t)  \tag{I}\\
\frac{d I(t)}{d t}=a S(t) I(t)-b I(t) \\
\frac{d R(t)}{d t}=b I(t)
\end{array}\right.
$$

where the given parameters $a, b \in \mathbb{R}^{+}$correspond to the rate of infection and recovery respectively, and $S(t), I(t)$ and $R(t)$ represent the number of susceptibles, infected and recovered as functions of time $t$, respectively. We can also include another term, namely $c S(t)$, which describes immunization effects through vaccination, resulting in the system

$$
\left\{\begin{align*}
\frac{d S(t)}{d t} & =-a S(t) I(t)-c S(t)  \tag{II}\\
\frac{d I(t)}{d t} & =a S(t) I(t)-b I(t) \\
\frac{d R(t)}{d t} & =b I(t)+c S(t)
\end{align*}\right.
$$

Equations (II) and (II) can be used to model an epidemic in a homogeneous population, meaning that the number of infectious individuals does not depend on their position in the domain. However, in applications usually this is not the case: one of the key elements of the investigation of the disease is the spatial spread of the infection in the population. This phenomenon is characterized by the effect of a single infectious individual, i.e., in what radius does an infectious individual have an effect on the susceptibles (or, in other words, one susceptible can be infected only by those whose are close to itself, namely those who are closer than a given distance). Such extensions of the Kermack-McKendrick model were first introduced by Kendall [67, 68] and the inclusion of such considerations result in a system of partial integro-differential equations.

It is also possible to extend Kendall's models with the introduction of latency period into the system, which means that there are some individuals who are infected, but not yet infectious - in this case class I includes not only the infectious, but also the infected ones. Such extensions result in a system of delay integro-differential equations.

In Chapters 1 and 2 we examine the properties of the space dependent and the delay system, respectively. In both sections we first prove that the equation has a unique solution, and then show that it possesses some biologically reasonable properties. Then some
numerical methods are constructed, and we show that (for a sufficiently chosen time step) the methods preserve the discrete versions of the properties of the continuous models. Sections 1.11 .6 collect the results of [103] and [104], while Section 1.7 summarizes the statements of [33]. Moreover, Chapter 2 collects the theorems of [105].

## Chapter 1

## Space dependent epidemic models

As it was mentioned in the introduction, the time-dependent functions in (II) represent the number of individuals in each class, but contain no information about their spatial distribution. Instead, one can replace these functions with space-dependent functions that describe the density of healthy, infectious and recovered species over some bounded domain $\Omega \subset \mathbb{R}^{d}\left(d \in \mathbb{N}^{+}\right)[103]$. From now on we consider a bounded domain in $\mathbb{R}^{2}$, hence system (II) is recast as

$$
\left\{\begin{align*}
\frac{\partial}{\partial t} S(t, x, y) & =-a S(t, x, y) I(t, x, y)-c S(t, x, y)  \tag{1.1}\\
\frac{\partial}{\partial t} I(t, x, y) & =a S(t, x, y) I(t, x, y)-b I(t, x, y) \\
\frac{\partial}{\partial t} R(t, x, y) & =b I(t, x, y)+c S(t, x, y)
\end{align*}\right.
$$

However, model (1.1) is still insufficient as it does not allow the disease to spread in the domain but only accounts for a point-wise infection. Spatial points do not interact with each other and infected individuals affect others only at their location.

In order to allow a realistic propagation of the infection, we assume that a healthy individual at a given point $(x, y)$ can be infected by ill species in a $\delta$-radius neighborhood around the point $(x, y)$ denoted by $B_{\delta}(x, y)$. Note though that infectious individuals are less likely to infect the healthy one at the center if the distance between them is bigger - because of this, we introduce a non-negative weighting function $G(r, \theta)$ describing the difference between of the effects of the different infectious species. Here the variable $r \in[0, \delta]$ describes the distance between the infectious individual and the healthy one at $(x, y)$, and $\theta \in[0,2 \pi)$ is the angle. We will also use the notations $\bar{x}(r, \theta)=x+r \cos (\theta)$ and $\bar{y}(r, \theta)=y+r \sin (\theta)$ for the coordinates of the infectious individuals. Now let us
define the weighting function $G(r, \theta)$ as

$$
G(r, \theta)= \begin{cases}g_{1}(r) g_{2}(\theta), & \text { if } r \leq \delta, \theta \in[0,2 \pi)  \tag{1.2}\\ 0, & \text { otherwise }\end{cases}
$$

Here we assumed that the right-hand side of $(1.2)$ is separable. The effect of the point $(x, y)$ depending on the distance from the center is described by $g_{1}(r)$ : a decreasing, non-negative function that is zero for values $r \geq \delta$ (since there is no effect outside $\left.B_{\delta}(x, y)\right)$. The bounded and non-negative function $g_{2}(\theta)$ characterizes the part of the effect depending on the angle, i.e., the direction in which the center is compared to point $(\bar{x}(r, \theta), \bar{y}(r, \theta))$. The case of constant function $g_{2}(\theta)$ is widely studied in 45] and 47], while such a non-constant function may be useful in the case of modeling the spread of diseases or a fire in a forest with a constant wind blowing in one direction. In both cases it is supposed that the function is periodic and bounded in the sense that $g_{2}(0)=\lim _{\theta \rightarrow 2 \pi} g_{2}(\theta)$. Note that later in Chapter 2 we assume that $g_{2}$ is a constant function. It is also worth mentioning that a weighting function depending on the location of the center point $(x, y)$ can also be introduced - such numerical experiments can be found in [104.

The nonlinear terms of the right-hand side of (1.1) describe the interactions between susceptible and infected species. We can now utilize 1.2 and replace the density of infected species in these nonlinear terms by

$$
\int_{0}^{\delta} \int_{0}^{2 \pi} G(r, \theta) I(t, \bar{x}(r, \theta), \bar{y}(r, \theta)) r \mathrm{~d} \theta \mathrm{~d} r
$$

where we used the fact that $G(x, y, r, \theta)=0$ outside the ball $B_{\delta}(x, y)$. Therefore, the model (1.1) can be extended as a system of integro-differential equations

$$
\left\{\begin{array}{l}
\frac{\partial S(t, x, y)}{\partial t}=-S(t, x, y) \int_{0}^{\delta} \int_{0}^{2 \pi} g_{1}(r) g_{2}(\theta) I(t, \bar{x}(r, \theta), \bar{y}(r, \theta)) r \mathrm{~d} \theta \mathrm{~d} r-c S(t, x, y)  \tag{1.3}\\
\frac{\partial I(t, x, y)}{\partial t}=S(t, x, y) \int_{0}^{\delta} \int_{0}^{2 \pi} g_{1}(r) g_{2}(\theta) I(t, \bar{x}(r, \theta), \bar{y}(r, \theta)) r \mathrm{~d} \theta \mathrm{~d} r-b I(t, x, y) \\
\frac{\partial R(t, x, y)}{\partial t}=b I(t, x, y)+c S(t, x, y)
\end{array}\right.
$$

Since we have no diffusion in our problem, we consider homogeneous Dirichlet conditions in the sense that we assume that there is no susceptible population outside of our domain. This means that we are going to assign a zero value to any point which lies outside of the domain $\Omega$ in which the problem is defined.

The structure of this chapter is as follows. In Section 1.1 we prove that the system (1.3) has a unique solution, and in Section 1.2 it is shown that this solution possesses some biologically reasonable properties. Since equation (1.3) can be rewritten as an integral equation, in Section 1.3 we present a numerical approximation of this integral equation. In Section 1.4 we discretize the system (1.3) in space, and then in Sections 1.5, 1.6 and 1.7 different time discretizations are examined, namely the method of Euler, Runge-Kutta methods and operator splitting techniques, respectively. The corresponding numerical experiments can be found at the end of each section.

### 1.1 Existence of the analytic solution

Analytic results for deterministic epidemic models have been studied by several authors, see for example, [8, 68, 106]. Such models lie in the larger class of reaction-diffusion problems and therefore one can obtain theoretical results by studying the more general problem. We prove the uniqueness of the solution for system (1.3) by following the work of Capasso and Fortunato [22].

We consider the following semilinear autonomous evolution problem

$$
\left\{\begin{align*}
\frac{d u}{d t}(t) & =-A u(t)+F(u(t))  \tag{1.4}\\
u_{0} & =u(0) \in D(A)
\end{align*}\right.
$$

where $A$ is a self-adjoint and positive-definite operator in a real Hilbert space $E$ with domain $D(A)$. Define $\lambda_{0}=\inf \sigma(A)$, where $\sigma(A)$ denotes the spectrum of $A$. Let us choose $u=\left(u_{1}, u_{2}\right)^{T} \in C^{1}\left(\left[0, t_{f}\right), D(A)\right)$ for some final time $t_{f} \in \mathbb{R}^{+}$and $E=L^{2}(\Omega) \times L^{2}(\Omega)$, where $\Omega$ is a bounded domain in $\mathbb{R}^{2}$, with a norm $\|\cdot\|$ defined by

$$
\begin{equation*}
\left\|\binom{u_{1}}{u_{2}}\right\|=\left(\left\|u_{1}\right\|_{L^{2}}^{2}+\left\|u_{2}\right\|_{L^{2}}^{2}\right)^{1 / 2} \tag{1.5}
\end{equation*}
$$

We also equip $D(A)$ with the norm

$$
\|u\|_{A}=\|A u\|, \quad u \in D(A) .
$$

Note that it is sufficient to consider only the first two equations in (1.3), since $R(t, x, y)$ can be obtained by using the fact that the sum $S(t, x, y)+I(t, x, y)+R(t, x, y)$ is constant in time for every point $(x, y)$. Hence, in view of problem (1.3), the linear operator $A$ is defined as

$$
A\binom{u_{1}}{u_{2}}=\left(\begin{array}{ll}
c & 0  \tag{1.6}\\
0 & b
\end{array}\right)\binom{u_{1}}{u_{2}}
$$

and $D(A)=E$. Since $b$ and $c$ are positive constants, it is easy to see that $A$ is a selfadjoint and positive-definite operator. Similarly, $F(u)$ consists of the nonlinear terms, and is defined as

$$
\begin{equation*}
F\binom{u_{1}}{u_{2}}=\binom{-u_{1} \mathcal{F}\left(u_{2}\right)}{u_{1} \mathcal{F}\left(u_{2}\right)} . \tag{1.7}
\end{equation*}
$$

The function $\mathcal{F}: L^{2}(\Omega) \rightarrow L^{2}(\Omega)$ contains the integral part of (1.3) and is given by

$$
\begin{equation*}
\mathcal{F}\left(u_{2}(x, y)\right)=\int_{0}^{\delta} \int_{0}^{2 \pi} g_{1}(r) g_{2}(\theta) u_{2}(\bar{x}(r, \theta), \bar{y}(r, \theta)) r \mathrm{~d} \theta \mathrm{~d} r . \tag{1.8}
\end{equation*}
$$

Note that in 1.3) $I(t, x, y)$ can be viewed as the map $I(t, x, y):\left[0, t_{f}\right) \rightarrow L^{2}(\Omega)$ acting as $t \longmapsto I_{t}(x, y) \in L^{2}(\Omega)$, in which $I_{t}(x, y):=I(t, x, y)$, thus the above definition makes sense.

The main result of this section is Theorem 1.1.1 stating that a unique solution of system (1.3) exists. Theorem 1.1.1 considers the system (1.4) as a generalization of (1.3) and its proof relies on the fact that the function $F$ in 1.7 is Lipschitz-continuous and bounded in $\|\cdot\|_{A}$. Therefore, we define the following conditions [22]:
$\left(A_{1}\right) F$ is locally Lipschitz-continuous from $D(A)$ to $D(A)$, i.e.,

$$
\|F(u)-F(v)\|_{A} \leq \zeta(d)\|u-v\|_{A}
$$

for all $u, v \in D(A)$ such that $d \geq 0$, and $\|u\|_{A} \leq d,\|v\|_{A} \leq d$. (Here $\zeta(d)$ is a constant depending on parameter $d$.)
$\left(A_{2}\right) F$ is bounded, i.e., there exist $\nu \geq 0$ and $\gamma \geq 0$ such that

$$
\|F(u)\|_{A} \leq \nu\|u\|_{A}^{1+\gamma}, \quad \forall u \in D(A) .
$$

We also denote the Lebesgue measure of $\Omega$ by $\mu(\Omega)$, and let

$$
\kappa_{1}=\max _{r \in(0, \delta)}\left\{g_{1}(r)\right\}, \quad \kappa_{2}=\max _{\theta \in[0,2 \pi)}\left\{g_{2}(\theta)\right\},
$$

and $\psi=\max \{b, c\} / \min \left\{b^{2}, c^{2}\right\}$.
Theorem 1.1.1. Consider the problem (1.3) and assume that conditions $\left(A_{1}\right)$ and ( $A_{2}$ ) hold. Then, a unique strong solution of system (1.3) exists on some interval $\left[0, t_{f}\right)$.

The proof of Theorem 1.1 .1 is a direct consequence of two main results by Capasso and Fortunato [22]. For clarity, we state these two theorems below.

Theorem 1.1.2. [22, Theorem 1.1] If assumption $\left(A_{1}\right)$ holds, then a unique strong solution in $D(A)$ of problem (1.4) exists in some interval $\left[0, t_{f}\right)$.

Theorem 1.1.3. [22, Theorem 1.3] Let us assume that $\left(A_{1}\right)$ and $\left(A_{2}\right)$ hold. Then for any $u_{0} \in \widetilde{K}$ a global strong solution in $D(A), u(t)$, of (1.4) exists. Moreover the zero solution is asymptotically stable in $\widetilde{K}$. Here

$$
\widetilde{K}= \begin{cases}\left\{u \in D(A) \mid\|u\|_{A}<\left(\lambda_{0} / \nu\right)^{1 / \gamma}\right\}, & \text { if } \gamma>0, \\ D(A), & \text { if } \gamma=0 \text { and } \lambda_{0}>\nu\end{cases}
$$

In 104 (by proving some simple lemmas) we show that the function $F$, as defined in (1.7), satisfies conditions $\left(A_{1}\right)$ and $\left(A_{2}\right)$.

We also get that the set $\widetilde{K}$ in Theorem 1.1 .3 can be computed by using that $D(A)=E$ and

$$
\left(\frac{\lambda_{0}}{\nu}\right)^{1 / \gamma}=\frac{\min \{b, c\}}{\sqrt{2} \psi \kappa_{1} \kappa_{2} \mu(\Omega)},
$$

where $b$ and $c$ are the diagonal elements of matrix $A$ in (1.6), $\lambda_{0}=\inf \sigma(A)$, and $\psi, \kappa_{1}, \kappa_{2}$ are as defined before. However, it is worth mentioning here that the zero convergence of the solution also holds for any non-negative initial condition, which will be a consequence of Theorem 1.2.1

In the next section we show that the unique solution (whose existence and uniqueness was stated in this section) has some biologically reasonable properties.

### 1.2 Properties of the analytic solution

When deriving a mathematical model to describe any physical, biological or chemical phenomena (here the spread of an epidemic in both space and time), it is essential that the real-life processes are being represented as accurately as possible. More precisely, not only the solution of the continuous model should possess properties resembling the real life processes, but numerical discretizations applied to such models should also preserve the qualitative properties of the original epidemic model.

The first, and perhaps most natural property is that the number of each species is non-negative at every time and point of the domain. Because in our models we use densities, this property can be formulated as follows:
$C_{1}$ : The densities $X(t, x, y)(X \in\{S, I, R\})$ are non-negative at every point $(x, y) \in \Omega$. Assuming that the births and natural deaths are the same (vital dynamics have no effect on the process), the total number of species (the sum of the species of each classes) should not increase nor decrease. Thus we have the following property:
$C_{2}$ : The sum $S(t, x, y)+I(t, x, y)+R(t, x, y)$ should be constant in time for every $(x, y) \in \Omega$, yielding

$$
\int_{\Omega} S(t, x, y)+I(t, x, y)+R(t, x, y) \mathrm{d} x \mathrm{~d} y=\text { const, } \quad \forall t .
$$

Another property concerns the number of susceptibles: since an individual gets to the recovered class after the infection, the number of susceptibles cannot increase in time.
$C_{3}$ : Function $S(t, x, y)$ is non-increasing in time at every $(x, y) \in \Omega$.
Similarly, the number of recovered species cannot decrease in time, thus:
$C_{4}$ : Function $R(t, x, y)$ is non-increasing in $t$ at every $(x, y) \in \Omega$.
As in the previous section, instead of proving the preservation of properties $C_{1}-C_{4}$ for the particular model (1.3), we can establish theoretical results for a more general system of equations. First, we state the following lemma, whose proof can be found in [104].

Lemma 1.2.1. The solution of (1.3) depends continuously on the right-hand side of the system of equations, meaning that if we consider the equation

$$
\left\{\begin{array}{l}
\frac{\partial S_{\varepsilon}(t, x, y)}{\partial t}=-S_{\varepsilon}(t, x, y) \int_{0}^{\delta} \int_{0}^{2 \pi} g_{1}(r) g_{2}(\theta) I_{\varepsilon}(t, \bar{x}(r, \theta), \bar{y}(r, \theta)) r \mathrm{~d} \theta \mathrm{~d} r-c S_{\varepsilon}(t, x, y)  \tag{1.9}\\
\frac{\partial I_{\varepsilon}(t, x, y)}{\partial t}=S_{\varepsilon}(t, x, y) \int_{0}^{\delta} \int_{0}^{2 \pi} g_{1}(r) g_{2}(\theta) I_{\varepsilon}(t, \bar{x}(r, \theta), \bar{y}(r, \theta)) r \mathrm{~d} \theta \mathrm{~d} r-b I_{\varepsilon}(t, x, y)+\varepsilon \\
\frac{\partial R_{\varepsilon}(t, x, y)}{\partial t}=b I_{\varepsilon}(t, x, y)+c S_{\varepsilon}(t, x, y)
\end{array}\right.
$$

in which $0 \leq \varepsilon \ll 1$ and consider a sequence of such $\varepsilon$ values denoted by $\left\{\varepsilon_{i}\right\}$, then as $\left\{\varepsilon_{i}\right\}$ tends to zero, the solutions $S_{\varepsilon_{i}}, I_{\varepsilon_{i}}$ and $R_{\varepsilon_{i}}$ converge in norm to the same limit regardless of the choice of the sequence $\left\{\varepsilon_{i}\right\}$.

The next theorem shows that the solution of (1.3) satisfies properties $C_{1}-C_{4}$.
Theorem 1.2.1. Suppose that the initial conditions of the system (1.3) are non-negative, i.e. $X(0, x, y) \geq 0, \forall(x, y) \in \Omega, X \in\{S, I, R\}$. In such case, the properties $C_{1}-C_{4}$ hold for the solution of (1.3) without any restriction on the time interval $t \in\left[0, t_{f}\right]$.

Proof. The proof consists of two parts: first we prove the required properties for a modified version of (1.3), and then by using Lemma 1.2 .1 we derive the statement of the theorem.

Consider equation (1.9) and let us suppose that the initial conditions assigned to the equation are all non-negative.

First, we would like to prove the non-negativity of $I_{\varepsilon}(t, x, y)$ by contradiction. Assume that the function takes negative values for some time $t$ at some point $(x, y) \in \Omega$. Let us denote by $t_{0}$ the last moment in time for which $I_{\varepsilon}(t, x, y)$ takes non-negative values, i.e.,

$$
t_{0}=\inf \left\{t \mid \exists(x, y) \in \Omega: I_{\varepsilon}(t, x, y)<0\right\} .
$$

By our assumptions, such $t_{0}$ value exists since $I_{\varepsilon}$ is continuous and the initial conditions are not negative, i.e., $I_{\varepsilon}(0, x, y) \geq 0$. Because of the continuity of $I_{\varepsilon}$ and the definition of $t_{0}$, there is a point $\left(x_{0}, y_{0}\right)$ for which $I_{\varepsilon}\left(t_{0}, x_{0}, y_{0}\right)=0$, and

$$
\begin{equation*}
\frac{\partial I_{\varepsilon}\left(t_{0}, x_{0}, y_{0}\right)}{\partial t} \leq 0 \tag{1.10}
\end{equation*}
$$

We know that all the values of $I_{\varepsilon}$ at $t_{0}$ inside $B_{\delta}\left(x_{0}, y_{0}\right)$ are non-negative by the definition of $t_{0}$, and $\mathcal{F}\left(I_{\varepsilon}\left(t_{0}, x_{0}, y_{0}\right)\right) \geq 0$ also holds.

However, if we observe the second equation in (1.9) at point $\left(t_{0}, x_{0}, y_{0}\right)$, we can see that the term $-b I_{\varepsilon}\left(t_{0}, x_{0}, y_{0}\right)$ is zero, so the term $S_{\varepsilon}\left(t_{0}, x_{0}, y_{0}\right) \mathcal{F}\left(I_{\varepsilon}\left(t_{0}, x_{0}, y_{0}\right)\right)$ must be negative for condition (1.10) to hold (since $\varepsilon$ is positive). We have already concluded that $\mathcal{F}\left(I_{\varepsilon}\left(t_{0}, x_{0}, y_{0}\right)\right) \geq 0$, so we need that $S_{\varepsilon}\left(t_{0}, x_{0}, y_{0}\right)<0$.

Now by dividing the first equation of (1.9) by $S_{\varepsilon}$ and integrating it with respect to time $t$ from 0 to $t_{0}$ yields

$$
\log \left(S_{\varepsilon}\left(t_{0}, x, y\right)\right)-\log \left(S_{\varepsilon}(0, x, y)\right)=-\int_{0}^{t_{0}} \mathcal{F}\left(I_{\varepsilon}\left(t_{0}, x, y\right)\right) \mathrm{d} t-c t_{0} .
$$

By reformulating, we get for $(x, y)=\left(x_{0}, y_{0}\right)$ that

$$
\begin{equation*}
S_{\varepsilon}\left(t_{0}, x_{0}, y_{0}\right)=S_{\varepsilon}\left(0, x_{0}, y_{0}\right) \exp \left(-\int_{0}^{t_{0}} \mathcal{F}\left(I_{\varepsilon}\left(t_{0}, x, y\right)\right) \mathrm{d} t-c t_{0}\right) \tag{1.11}
\end{equation*}
$$

Therefore $S_{\varepsilon}\left(t_{0}, x_{0}, y_{0}\right)$ is non-negative, so we get a contradiction.
As a result, $I_{\varepsilon}(t, x, y) \geq 0$ for every $t \in\left[0, t_{f}\right]$ and $(x, y) \in \Omega$. Consequently, since $R_{\varepsilon}(0, x, y)$ is non-negative we get that $R_{\varepsilon}(t, x, y)$ is a non-decreasing and a non-negative function. Note also that the calculations resulting in formula (1.11) are also true for any time $t$ and point $(x, y) \in \Omega$, meaning that $S_{\varepsilon}$ is also non-negative, and since $\mathcal{F}\left(I_{\varepsilon}\left(t_{0}, x, y\right)\right)$ is non-negative, we also get the non-increasing property from the first equation of (1.9). Hence, we proved that the solution of (1.9) satisfies $C_{1}-C_{4}$.

Finally, we also know that by Lemma 1.2.1,

$$
\left.\lim _{\varepsilon \rightarrow 0} X_{\varepsilon}(t, x, y)\right|_{t \in\left[0, t_{f}\right]}-\left.X(t, x, y)\right|_{t \in\left[0, t_{f}\right]}=0
$$

holds for every $X \in\{S, I, R\}$. Therefore, properties $C_{1}-C_{4}$ are also satisfied by the solution of system (1.3).

The following corollary is a consequence of the previously proved properties and can be proved easily.

Corollary 1.2.1. Functions $S(t, x, y)$ and $I(t, x, y)$ in the solution of (1.3) tend to zero as $t \rightarrow \infty$.

Due to the complicated form of the equations in (1.3) one can suspect that no analytic
solution can be derived for this system. Because of this, we are going to use numerical methods to approximate the solution of these equations. However, the analytic solution of the original SIR model (I) has been described in the papers by Harko et al. [59] and Miller [80, 81]. Thus, we can get similar results applying their observations to our modified model (1.3).

The analytic solution of system (1.3) can be written as

$$
\left\{\begin{array}{l}
S(t, x, y)=S(0, x, y) e^{-\phi(t, x, y)-c t},  \tag{1.12}\\
I(t, x, y)=M_{0}(x, y)-S(t, x, y)-R(t, x, y) \\
R(t, x, y)=R(0, x, y)+b \int_{0}^{t} I(s, x, y) \mathrm{d} s+c \int_{0}^{t} S(s, x, y) \mathrm{d} s,
\end{array}\right.
$$

where we use the notations

$$
\begin{aligned}
& M_{0}(x, y)=S(0, x, y)+I(0, x, y)+R(0, x, y), \\
& \phi(t, x, y)=\int_{0}^{t} \mathcal{F}(I(s, x, y)) \mathrm{d} s
\end{aligned}
$$

and $\mathcal{F}$ is given by (1.8).
It is evident that in $(1.12)$, the values of the functions at a given time $t^{*}$ can only be computed if all the values of this function are known at all the points in the interval $\left[0, t^{*}\right)$. Consequently, these formulas are not useful in practice, since (1.12) is an implicit system in the solutions $S(t, x, y), I(t, x, y)$ and $R(t, x, y)$. Later (see Table 1.2 in Section 1.6.3), an approximation of the solution of 1.12 will be compared to the numerical solution computed using the first-order forward Euler scheme.

Since the values of the functions in (1.12) cannot be calculated directly, numerical methods are needed to approximate them. We can take two possible paths:

1. approximate the values of $\phi(t, x, y)$ and the integrals in the third equation of (1.12) by numerical integration; or
2. approximate the solution of the original equation (1.3) by a numerical method.

The first approach is discussed in Section 1.3, while Sections 1.4 1.7 consider the second case. We focus on the convergence rate of our numerical methods, and ensure that qualitative properties $C_{1}-C_{4}$ of the analytic solution are preserved by the numerical method.

For that, a discrete analogue of conditions $C_{1}-C_{4}$ (later denoted by $D_{1}-D_{4}$ ) is required; see Section 1.5

### 1.3 Numerical approximation of the integral solution

As noted before, if we would like to use the solution 1.12 then we have to approximate the involved integrals. This can be achieved by partitioning the time interval $\left[0, t_{f}\right]$ into uniform spaced sections by using a constant time step $\tau$. With this approach, the integrals can be approximated by a left Riemann sum, meaning that if $t \in\left[t_{i}, t_{i+1}\right)$, then we approximate the densities $X(t, x, y)$ by $X\left(t_{i}, x, y\right)(X \in\{S, I, R\})$. Therefore, for any integer $1 \leq n \leq \mathcal{N}$ such that $t_{f}=\tau \mathcal{N}$, the integral of $X(t, x, y)$ can be approximated by

$$
\int_{0}^{n \tau} X(s, x, y) \mathrm{d} s \approx \tau \sum_{k=0}^{n-1} X(k \tau, x, y)
$$

Note that later we will also use a right Riemann sum, in which the value $X(t, x, y)$ is approximated by $X\left(t_{i+1}, x, y\right)$ - in this case

$$
\int_{0}^{n \tau} X(s, x, y) \mathrm{d} s \approx \tau \sum_{k=1}^{n} X(k \tau, x, y)
$$

An important observation is that the integral equations (1.12) can be rewritten in a recursive form

$$
\left\{\begin{array}{l}
S(n \tau, x, y)=S((n-1) \tau, x, y) \exp \left(-\int_{(n-1) \tau}^{n \tau} \mathcal{F}(I(s, x, y)) \mathrm{d} s-c \tau\right)  \tag{1.13}\\
R(n \tau, x, y)=R((n-1) \tau, x, y)+b \int_{(n-1) \tau}^{n \tau} I(s, x, y) \mathrm{d} s+c \int_{(n-1) \tau}^{n \tau} S(s, x, y) \mathrm{d} s \\
I(n \tau, x, y)=M_{0}(x, y)-S(n \tau, x, y)-R(n \tau, x, y)
\end{array}\right.
$$

Let $X^{n}(x, y) \approx X(n \tau, x, y), X \in\{S, I, R\}$, and define $\mathcal{F}^{n}=\mathcal{F}\left(I^{n}\right)$. Using the approximations

$$
\begin{aligned}
\tau \mathcal{F}^{n-1} & \approx \int_{(n-1) \tau}^{n \tau} \mathcal{F}(I(s, x, y)) \mathrm{d} s \\
\tau I^{n-1} & \approx \int_{(n-1) \tau}^{n \tau} I(s, x, y) \mathrm{d} s \\
\tau S^{n} & \approx \int_{(n-1) \tau}^{n \tau} S(s, x, y) \mathrm{d} s
\end{aligned}
$$

(note that in the first two were are using left, while in the third one a right Riemann sum), we get an approximating scheme for (1.12), given by

$$
\left\{\begin{align*}
S^{n} & =S^{n-1} e^{-\tau \mathcal{F}^{n-1}-c \tau}  \tag{1.14a}\\
R^{n} & =R^{n-1}+b \tau I^{n-1}+c \tau S^{n} \\
I^{n} & =\left(S^{n-1}+I^{n-1}+R^{n-1}\right)-S^{n}-R^{n}
\end{align*}\right.
$$

Note that in this case, the order of the equations in (1.14) is important as estimates at time $t_{n}=n \tau$ are used to update the rest of the components of the solution. Also, it is worth mentioning that the choice for left or right Riemann sums was arbitrary - one can also use the same type in all three cases.

Theorem 1.3.1. Consider the solution $X^{n}(x, y), X \in\{S, I, R\}$ of scheme (1.14) on the time interval $\left[0, t_{f}\right]$, where $1 \leq n \leq \mathcal{N}$. Let $\mathcal{N}$ be the total number of steps such that $t_{f}=\tau \mathcal{N}$, where $\tau$ denotes the time step. If the step-size restriction $0<\tau \leq 1 / b$ holds, then the solution of (1.14) satisfies properties $D_{1}-D_{4}$ at times $t_{n}=n \tau, 1 \leq n \leq \mathcal{N}$.

For the proof, see [104]
Remark 1.3.1. Using left Riemann sums to approximate the integrals in (1.13) results in local errors of order $\mathcal{O}\left(\tau^{2}\right)$. Therefore, the solution of (1.13) can only be first order accurate.

In Section 1.4 we discretize (1.3) by first using a numerical approximation of the integral on the right hand side of the system, and then in Sections $1.5,1.6$ and 1.7 apply a time integration method. This approach results in numerical schemes that are (in most cases) high order accurate, both in space and time, meaning that they are more favorable than the method described in this section. However, it is worth mentioning here that it
is also possible to use higher order cubatures for the approximation of the integrals in the integral equations, but these methods are not discussed in this work.

### 1.4 Spatial discretization

It is evident that the key element of the numerical solution of problem (1.3) is the approximation of the double integrals in $\mathcal{F}(I(t, x, y))$. This can be done in two different ways. The first approach is to approximate the function $I(t, \bar{x}(r, \theta), \bar{y}(r, \theta))$ by a Taylor expansion, and then integrating the Taylor polynomial as it was described in [45, 47].

It turned out that the spatial Taylor expansion of the term $I(t, \bar{x}(r, \theta), \bar{y}(r, \theta))$ around the point $(x, y)$ results in a method which can be used successfully. However, this method cannot be used efficiently in the general case, for the set of functions

$$
\mathcal{H}:=\left\{c_{1} \sin \theta+c_{2} \cos \theta+c_{3}\right\} .
$$

This is a useful choice in biological terms for a set from which the function $g_{2}(\theta)$ would be chosen, since it contains smooth functions which are also periodic. One can also think of the functions from $\mathcal{H}$ as a first order trigonometric approximation of some real-life but much more complicated function. Later the function $g_{2}(\theta)=\beta \sin (\theta+\alpha)+\beta$ is used ( $\alpha$, $\beta \geq 0$ ), which function is also from the set $\mathcal{H}$.

For functions from $\mathcal{H}$, we cannot use the fact that the first order terms in the Taylor expansion are zero (which was used in [45, 47]). The reason behind this is the fact that in the non-constant case (when $c_{1}^{2}+c_{2}^{2} \neq 0$ ) these first order terms cannot be zero at the same time, resulting in a more complicated numerical model.

The other approach is to use a combination of interpolation and numerical integration (by using cubature formulas) to obtain an approximation of $\mathcal{F}(I(t, x, y))$.

We consider two-dimensional cubature formulas on the disc of radius $\delta$ with positive coefficients. Denote by $\mathcal{Q}(x, y)$ the set of cubature nodes in the disk $B_{\delta}(x, y)$ parametrized by polar coordinates, i.e.,

$$
\mathcal{Q}(x, y):=\left\{\left(x_{i j}, y_{i j}\right)=\left(x+r_{i} \cos \left(\theta_{j}\right), y+r_{i} \sin \left(\theta_{j}\right)\right) \in B_{\delta}(x, y), i \in \mathcal{I}, j \in \mathcal{J}\right\}
$$

where $r_{i}$ denotes the distance from center point $(x, y), \theta_{j}$ is the angle, and $\mathcal{I}$ and $\mathcal{J}$ are
the set of indices of cubature nodes. Using numerical integration, we get the system

$$
\left\{\begin{array}{l}
\frac{\partial S(t, x, y)}{\partial t}=-S(t, x, y) T(t, \mathcal{Q}(x, y))-c S(t, x, y)  \tag{1.15}\\
\frac{\partial I(t, x, y)}{\partial t}=S(t, x, y) T(t, \mathcal{Q}(x, y))-b I(t, x, y) \\
\frac{\partial R(t, x, y)}{\partial t}=b I(t, x, y)+c S(t, x, y)
\end{array}\right.
$$

where

$$
T(t, \mathcal{Q}(x, y))=\sum_{\left(x_{i j}, y_{i j}\right) \in \mathcal{Q}(x, y)} w_{i, j} g_{1}\left(r_{i}\right) g_{2}\left(\theta_{j}\right) I\left(t, x+r_{i} \cos \left(\theta_{j}\right), y+r_{i} \sin \left(\theta_{j}\right)\right),
$$

and $w_{i, j}>0$ are the weights of the cubature formula.
Remark 1.4.1. Note that similar arguments as the proof of Theorem 1.2.1 can be applied to system 1.15; hence, the properties $C_{1}-C_{4}$ hold without any restrictions for the analytic solution of this system. Moreover, it can be easily shown that $T(t, \mathcal{Q}(x, y))$ satisfies properties $\left(A_{1}\right)$ and $\left(A_{2}\right)$, by similar arguments that can be applied to the original system. As a result, system (1.15) admits a unique strong solution.

### 1.4.1 The semi-discretized system

In this section we would like to solve (1.15) numerically. The first step is to discretize the problem in space.

Let us suppose that we would like to solve our problem on a rectangle-shaped domain, namely on $\Omega:=\left[0, \mathcal{L}_{1}\right] \times\left[0, \mathcal{L}_{2}\right]$. For our numerical solutions we will discretize this domain by using a spatial grid

$$
\mathcal{G}:=\left\{\left(x_{k}, y_{\ell}\right) \in \Omega \mid 1 \leq k \leq P_{1}, 1 \leq \ell \leq P_{2}\right\},
$$

which consists of $P_{1} \times P_{2}$ points with spatial step sizes $h_{1}$ and $h_{2}$, and approximate the continuous solutions by a vector of the values at the grid points. After this semi-
discretization, we get the following set of equations

$$
\left\{\begin{array}{l}
\frac{d S_{k, \ell}(t)}{d t}=-S_{k, \ell}(t) T_{k, \ell}\left(t, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right)-c S_{k, \ell}(t)  \tag{1.16}\\
\frac{d I_{k, \ell}(t)}{d t}=S_{k, \ell}(t) T_{k, \ell}\left(t, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right)-b I_{k, \ell}(t) \\
\frac{d R_{k, \ell}(t)}{d t}=b I_{k, \ell}(t)+c S_{k, \ell}(t)
\end{array}\right.
$$

where $X_{k, \ell}(t)(X \in\{S, I, R\})$ denotes the approximation of the function at gridpoint $\left(x_{k}, y_{\ell}\right)$. The approximation of $\mathcal{F}\left(t, x_{k}, y_{\ell}\right)$ is denoted by $T_{k, \ell}\left(t, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right)$ and defined as

$$
\begin{equation*}
T_{k, \ell}\left(t, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right):=\sum_{\left(\bar{x}_{k}, \bar{y}_{l}\right) \in \mathcal{Q}\left(x_{k}, y_{\ell}\right)} w_{i, j} g_{1}\left(r_{i}\right) g_{2}\left(\theta_{j}\right) z\left(t, \bar{x}_{k}, \bar{y}_{l}\right), \tag{1.17}
\end{equation*}
$$

where $\bar{x}_{k}=x_{k}+r_{i} \cos \left(\theta_{j}\right), \bar{y}_{l}=y_{\ell}+r_{i} \sin \left(\theta_{j}\right)$ and $z$ is an interpolation of $I\left(t, \bar{x}_{k}, \bar{y}_{l}\right)$ by using a positivity preserving interpolation (e.g. bilinear interpolation) with the nearest known $I_{k, \ell}$ values and positive coefficients. The reason for this is that the points ( $\bar{x}_{k}, \bar{y}_{l}$ ) might not be included in $\mathcal{G}$; in such case there are no $I_{k, \ell}$ values assigned to them.

Theorem 1.4.1. A unique strong solution for system (1.16) exists, for which properties $C_{1}-C_{4}$ hold locally at a given point $\left(x_{k}, y_{\ell}\right)$.

Proof. The proof of existence and uniqueness comes from the Lipschitz continuity and boundness of the right hand side, which can be proved similarly as in the case of the original system (1.3). Properties $C_{1}-C_{4}$ can be proved in a similar manner as in Theorem 1.2.1.

Remark 1.4.2. The previous theorem can be stated not only in the case of bilinear interpolation, but also for any positivity preserving interpolation.

The next theorem characterizes the accuracy of interpolation and cubature techniques of system 1.16).

Theorem 1.4.2. Suppose that a cubature rule approximates the integral 1.8) to order p, i.e.,

$$
\begin{equation*}
\|\mathcal{F}(I(t, x, y))-T(t, \mathcal{Q}(x, y))\|_{L^{2}}=\mathcal{O}\left(\delta^{p}\right), \tag{1.18}
\end{equation*}
$$

where $\delta$ is the radius of the disk in which the integration takes place. Let us suppose that the (positivity preserving) spatial interpolation $z$ approximates the values of I to order $q$,
i.e.,

$$
\begin{equation*}
\|I(t, x, y)-z(t, x, y)\|_{L^{2}}=\mathcal{O}\left(h^{q}\right), \tag{1.19}
\end{equation*}
$$

where $h=\min \left\{h_{1}, h_{2}\right\}$ is the minimum of the spatial step sizes. Then if $\tilde{u}$ is the vector of the solution of (1.3) evaluated at the grid points of $\mathcal{G}$ and $\tilde{v}$ is the vector of the solution of (1.16), it follows that

$$
\|\tilde{u}-\tilde{v}\|_{L^{2}, d}=\mathcal{O}\left(\delta^{p}\right)+\mathcal{O}\left(h^{q}\right) .
$$

in which $\|\cdot\|_{L^{2}, d}$ is the discrete $L^{2}$-norm.
For the proof, see 104 .
A natural question arises: what is the best type of cubature and interpolation for solving the system (1.16)? In the rest of the section we describe two numerical integration procedures and also discuss suitable interpolation techniques.

### 1.4.1.1 Elhay-Kautsky cubature

One can use a direct cubature rule on the general disk, see for example [38, 97]. In such cases the integral of a function $f(x, y)$ over the disk with radius $\delta$ can be approximated by

$$
\begin{equation*}
Q(f)=\pi \delta^{2} \sum_{i=1}^{N_{r} \cdot N_{\theta}} w_{i} f\left(x_{i}, y_{i}\right)=\pi \delta^{2} \sum_{i=1}^{N_{r}} \sum_{j=1}^{N_{\theta}} \widetilde{w}_{i} f\left(r_{i} \cos \left(\theta_{j}\right), r_{i} \sin \left(\theta_{j}\right)\right), \tag{1.20}
\end{equation*}
$$

where $N_{r}$ is the number of radial nodes, $N_{\theta}$ is the number of equally spaced angles, and $w_{i}$ and $\widetilde{w}_{i}$ are weights in the $[0,1]$ interval. We use $N_{\theta}=2 N_{r}$ to have a cubature rule that is equally powerful in both $r$ and $\theta$. The weights and cubature nodes are calculated by a modification of the Elhay-Kautsky Legendre quadrature method [42, 65, 78]. The top panel of Figure 1.1 shows the distribution of cubature nodes for $N_{r} \in\{3,6,12\}$. The Elhay-Kautsky cubature results in nodes that are evenly spaced in the $\theta$ direction.

### 1.4.1.2 Gauss-Legendre quadrature

Alternatively, we can transform the disk into a square, and then use a one-dimensional Gauss-Legendre rule to approximate the integral. First, we transform the disk with radius
$\delta$ to the rectangle $[0, \delta] \times[0,2 \pi]$ on the $r-\theta$ plane. Next, the rectangle $[0, \delta] \times[0,2 \pi]$ is mapped to $[0,1] \times[0,1]$ on the $\xi-\eta$ plane by using the linear transformation

$$
r=\delta \xi, \quad \theta=2 \pi \eta
$$

that has a Jacobian $2 \pi \delta$. Using these transformations, the original integral

$$
\int_{0}^{\delta} \int_{0}^{2 \pi} f(r \cos (\theta), r \sin (\theta)) r \mathrm{~d} \theta \mathrm{~d} r
$$

takes the form

$$
\begin{equation*}
\int_{0}^{1} \int_{0}^{1} f(\delta \xi \cos (2 \pi \eta), \delta \xi \sin (2 \pi \eta)) \delta \xi 2 \pi \delta \mathrm{~d} \eta \mathrm{~d} \xi \tag{1.21}
\end{equation*}
$$

There are several approaches for computing multiple integrals based on numerical integration of one-dimensional integrals. In this paper, we use the Gauss-Legendre quadrature rule on the unit interval [107]; other options include generalized Gaussian quadrature rules as described in [75].

The integral (1.21) can be approximated by

$$
\begin{equation*}
Q(f)=\sum_{i=1}^{N_{\xi}} \sum_{j=1}^{N_{\eta}} w_{i} w_{j} 2 \pi \delta^{2} \xi_{i} f\left(\delta \xi_{i} \cos \left(2 \pi \eta_{j}\right), \delta \xi_{i} \sin \left(2 \pi \eta_{j}\right)\right)=\sum_{m=1}^{N_{\xi} \cdot N_{\eta}} \widetilde{w}_{m} f\left(x_{m}, y_{m}\right) \tag{1.22}
\end{equation*}
$$

where $\xi_{i}$ and $\eta_{i}$ are the $i$ th cubature nodes corresponding to the Gauss-Legendre quadrature with weights $w_{i}$. The number of cubature nodes in the $\xi$ and $\eta$ direction are denoted by $N_{\xi}$ and $N_{\eta}$, respectively, and we let $x_{m}=\delta \xi_{i} \cos \left(2 \pi \eta_{j}\right), y_{m}=\delta \xi_{i} \sin \left(2 \pi \eta_{j}\right)$ and $\widetilde{w}_{m}=w_{i} w_{j} 2 \pi \delta^{2} \xi_{i}$. The distribution of the cubature nodes in the unit disk is not uniform as with the Elhay-Kautsky cubature and can be seen in the bottom panel of Figure 1.1. For a fair comparison we use $N_{\eta}=2 N_{\xi}$. Experimental results reveal that the ElhayKautsky cubature (1.20) performs better in cases the interpolated function $f(x, y)$ is a bivariate polynomial, whereas the Gauss-Legendre quadrature (1.22) or the generalized Gaussian quadrature rule (see [75]) when $f(x, y)$ is an arbitrary nonlinear function.

In order to determine which cubature rule performs better for the system (1.16), we perform a convergence test by applying the cubature formulas 1.20 and 1.22 to the


Figure 1.1: Top panel: The distribution of cubature nodes $\left(N_{r} \times N_{\theta}\right)$ in the unit disk using the Elhay-Kautsky cubature rule. Bottom panel: The distribution of cubature nodes $\left(N_{\xi} \times N_{\eta}\right)$ in the unit disk using the Gauss-Legendre quadrature rule.
function $g_{1}(r) g_{2}(\theta) I_{0}(r, \theta) r$, where

$$
g_{1}(r)=100(-r+\delta), \quad g_{2}(\theta)=\sin (\theta)+1,
$$

and

$$
I_{0}(r, \theta)=\frac{100}{2 \pi \sigma^{2}} \exp \left(-\frac{r^{2}}{2 \sigma^{2}}\right)
$$

is a Gaussian distribution with deviation $\sigma$ and centered at zero. This resembles the initial conditions for $I$ at the origin, as we will use later in Section 1.6.3

The exact solution of the integral over a disk of radius $\delta$ is given by

$$
\begin{equation*}
\int_{0}^{\delta} \int_{0}^{2 \pi} g_{1}(r) g_{2}(\theta) I_{0} r \mathrm{~d} \theta \mathrm{~d} r=5000\left(2 \delta-\sqrt{2 \pi} \sigma \operatorname{erf}\left(\frac{\delta}{\sqrt{2} \sigma}\right)\right) \tag{1.23}
\end{equation*}
$$

where $\operatorname{erf}(x)$ is the Gauss error function [6, 57]. Figure 1.2 shows the convergence of the two cubature rules over the disk of radius $\delta$, as $\delta$ goes to zero ( $\sigma=1 / 10$ ). We observe that the Gauss-Legendre quadrature (1.22) gives much smaller errors (close to machine precision) when more than $12 \times 24$ nodes are used, compared to the Elhay-Kautsky cubature 1.20 which is third-order accurate.


Figure 1.2: Numerical integration errors of cubature formulas (1.20) and (1.22) applied to the integral in 1.23). The colored curves correspond to different choices of cubature nodes in the $\delta$-radius disk.

The performance of the cubature formulas depends also on the choice and accuracy of interpolation. As mentioned before, bilinear interpolation can be used since it preserves the non-negativity of the interpolant. One possibility is to use higher order interpolations, like cubic or spline, but in these cases the preservation of the required properties cannot be guaranteed. However, numerical experiments show that piecewise cubic spline interpolation results in a positive interpolant for sufficiently fine spatial grid. A better choice is the use of a shape-preserving interpolation, to ensure that negative values are not generated and the interpolant of $I\left(t, \bar{x}_{k}, \bar{y}_{l}\right)$ in (1.17) is bounded by $\max _{k, \ell}\left\{S_{k, \ell}+I_{k, \ell}+R_{k, \ell}\right\}$ for every point $\left(x_{k}, y_{\ell}\right)$. This can be accomplished by a monotone interpolation that uses piecewise cubic Hermite interpolating polynomials [41, 49]. In MATLAB (version R2020b) the relevant function is called pchip but is only available for one-dimensional problems (in Octave it is also avaible in two dimensions). Extensions to bivariate shape-preserving interpolation have been studied in [23, 24, 50]; however, this topic goes beyond the purposes of this work. Another choice is the modified Akima piecewise cubic Hermite interpolation, makima. Numerical experiments demonstrate good performance as it avoids
overshoots when more than two consecutive nodes are constant [2, 3], and hence preserves non-negativity in areas where $I\left(t, \bar{x}_{k}, \bar{y}_{l}\right)$ is close to zero.

### 1.5 Discretization in time: the method of Euler

Let us apply the explicit Euler method to system (1.16) on the interval $\left[0, t_{f}\right]$, and choose an adaptive time step $\tau_{n}>0$ such that $t_{n}=t_{n-1}+\tau_{n}, n \geq 1$. After the full discretization we get the set of algebraic equations

$$
\left\{\begin{array}{l}
S^{n}=S^{n-1}-\tau_{n} S^{n-1} \circ T^{n-1}-c \tau_{n} S^{n-1}  \tag{1.24a}\\
I^{n}=I^{n-1}+\tau_{n} S^{n-1} \circ T^{n-1}-b \tau_{n} I^{n-1} \\
R^{n}=R^{n-1}+b \tau_{n} I^{n-1}+c \tau_{n} S^{n-1}
\end{array}\right.
$$

Here, the operator o denotes the element-by-element or Hadamard product of matrices.
Now we examine the bounds of time step $\tau_{n}$ such that the method (1.24) gives solutions which are qualitatively adequate and satisfy conditions $D_{1}-D_{4}$.

Theorem 1.5.1. Consider the numerical solution (1.24) obtained by forward Euler method applied to (1.16) with non-negative initial data. Then, the solution satisfies property $D_{2}$ without any step-size restrictions. Moreover, properties $D_{1}, D_{3}$ and $D_{4}$ hold if the time step satisfies

$$
\begin{equation*}
\tau_{n} \leq \min \left\{\frac{1}{\max _{k, \ell}\left\{T_{k, \ell}^{n-1}\right\}+c}, \frac{1}{b}\right\} \tag{1.25}
\end{equation*}
$$

where

$$
\begin{equation*}
T_{k, \ell}^{n-1}=\sum_{\left(\bar{x}_{k}, \bar{y}_{l}\right) \in \mathcal{Q}\left(x_{k}, y_{\ell}\right)} w_{i, j} g_{1}\left(r_{i}\right) g_{2}\left(\theta_{j}\right) z^{n-1}\left(\bar{x}_{k}, \bar{y}_{l}\right) \tag{1.26}
\end{equation*}
$$

is an approximation of (1.17) at point $\left(x_{k}, y_{\ell}\right) \in \mathcal{G}$.
For the proof, see [104].
A drawback of the time-step restriction (1.25) is that it depends on the solution at the previous step. This has important complications for higher order methods as we will see in Section 1.6. For any multistage method the adaptive time step bound 1.25
depends not only on the previous solution, but also on the internal stage approximations. Consequently, an adaptive time-step restriction based on (1.25) cannot be the same for all stages of a Runge-Kutta method; instead it needs to be recalculated at every stage to guarantee that conditions $D_{1}-D_{4}$ hold. Therefore, such bound has no practical use because it is prone to rejected steps and will likely tend to zero.

A remedy is to use a constant time step that is less strict than (1.25), but still guarantee that $\tau \leq 1 /\left(T_{k, \ell}^{n-1}+c\right)$ holds for all $1 \leq k \leq P_{1}, 1 \leq l \leq P_{2}$ and at every step $n$. At a given point $\left(x_{k}, y_{\ell}\right) \in \mathcal{G}$ the weights and cubature nodes in $B_{\delta}\left(x_{k}, y_{\ell}\right)$ are the same regardless of the location of $\left(x_{k}, y_{\ell}\right)$ in the domain. Therefore, we can find an upper bound for each element of the matrix $T^{n-1}$ in (1.26). Let

$$
\begin{equation*}
\widehat{T}:=\sum_{\left(\bar{x}_{k}, \bar{y}_{l}\right) \in \mathcal{Q}\left(x_{k}, y_{\ell}\right)} w_{i, j} g_{1}\left(r_{i}\right) g_{2}\left(\theta_{j}\right) \widetilde{m} \tag{1.27}
\end{equation*}
$$

where

$$
\begin{equation*}
\widetilde{m}=\max _{\left(x_{k}, y_{\ell}\right) \in \mathcal{G}}\left\{S\left(0, x_{k}, y_{\ell}\right)+I\left(0, x_{k}, y_{\ell}\right)+R\left(0, x_{k}, y_{\ell}\right)\right\} . \tag{1.28}
\end{equation*}
$$

Since $T_{k, \ell}^{n-1} \leq \widehat{T}$ for all $1 \leq k \leq P_{1}, 1 \leq l \leq P_{2}$ then if

$$
\begin{equation*}
\widehat{\tau}:=\min \left\{\frac{1}{\widehat{T}+c}, \frac{1}{b}\right\} \tag{1.29}
\end{equation*}
$$

then the condition

$$
\widehat{\tau} \leq \min \left\{\frac{1}{\max _{k, \ell}\left\{T_{k, \ell}^{n-1}\right\}+c}, \frac{1}{b}\right\}
$$

holds at every step $n$. Moreover, $\widehat{T} \leq \widetilde{w} \kappa^{2} \widetilde{m} N$, where

$$
\kappa=\max \left\{\kappa_{1}, \kappa_{2}\right\}=\max \left\{\max _{r \in(0, \delta)}\left\{g_{1}(r)\right\}, \max _{\theta \in[0,2 \pi)}\left\{g_{2}(\theta)\right\}\right\},
$$

$\widetilde{w}=\max _{i, j}\left\{w_{i, j}\right\}$, and $N$ is the number of the cubature nodes in $\mathcal{Q}\left(x_{k}, y_{\ell}\right)$. Hence, the time step 1.29 is larger than the rather pessimistic time step

$$
\begin{equation*}
\widetilde{\tau}:=\min \left\{\frac{1}{\widetilde{w} \kappa^{2} \widetilde{m} N+c}, \frac{1}{b}\right\} \tag{1.30}
\end{equation*}
$$

proposed in [103, Theorem 2]. Numerical experiments show that $\widehat{\tau}$ is very close to the theoretical bound in 1.25 , and thus a relatively small increase of time step beyond the bound (1.29) may produce qualitatively bad solutions which violate one of the conditions $D_{1}-D_{4}$ (see Section 1.6.3).

It is also possible to consider an implicit-explicit (IMEX) scheme, and apply it to the system (1.16), which results in the scheme

$$
\left\{\begin{array}{l}
S^{n+1}=S^{n}-\tau_{n} S^{n} \circ T^{n}-c \tau_{n} S^{n+1}  \tag{1.31}\\
I^{n+1}=I^{n}+\tau_{n} S^{n} \circ T^{n}-b \tau_{n} I^{n+1} \\
R^{n+1}=R^{n}+b \tau_{n} I^{n+1}+c \tau_{n} S^{n+1}
\end{array}\right.
$$

in which $S^{n}, I^{n}, R^{n}$ denote the numerical solutions at $t_{n}$ as before.
Theorem 1.5.2. Property $D_{2}$ holds without restrictions, and if the time step satisfies

$$
\begin{equation*}
\tau_{n} \leq \frac{1}{\max _{k, \ell}\left\{T_{k, \ell}^{n-1}\right\}} \tag{1.32}
\end{equation*}
$$

then properties $D_{1}, D_{3}$ and $D_{4}$ also hold.

Proof. The proof is the same as the one of Theorem 1.5.1, except that in this case $\tau$ does not have to be smaller than $1 / b$.

It is evident that the bound $(1.32)$ is a less strict one than 1.25 , so the use of this implicit-explicit scheme enables the use of larger time steps. However, since these are only first order methods, their use is not recommended - in the next section higher order methods are considered.

### 1.6 Discretization in time: Runge-Kutta methods

The method of Euler (either the explicit or explicit-implicit one) is only first-order accurate; hence, we would like to obtain time step restrictions for higher order Runge-Kutta methods. Note that the spatial discretizations discussed in Section 1.4 can be chosen so that errors from cubature formulas and interpolation are very small; therefore, it is substantial to have a high-order accurate time integration method.

### 1.6.1 Strong stability preserving Runge-Kutta methods

Consider an initial value problem, that usually results from semi-discretization of partial differential equations:

$$
\left\{\begin{array}{l}
U^{\prime}(t)=\mathcal{P}(U(t)),  \tag{1.33}\\
U(0)=U_{0}
\end{array}\right.
$$

in which $\mathcal{P}$ is the operator describing the discretization in space, and $U_{0}$ is a given function. Then, if we apply the Runge-Kutta method given in the usual Butcher form [21] to problem (1.33), the scheme we get is

$$
\begin{align*}
U^{(i)} & =U^{n-1}+\tau \sum_{j=1}^{n} a_{i j} \mathcal{P}\left(U^{(j)}\right), \quad 1 \leq i \leq m  \tag{1.34}\\
U^{n} & =U^{n-1}+\tau \sum_{j=1}^{n} b_{j} \mathcal{P}\left(U^{(j)}\right),
\end{align*}
$$

in which $U^{n}=\left(U_{1}^{n}, \ldots, U_{N}^{n}\right)$ is the numerical solution at time $t_{n}$ and $U^{(i)}=\left(U_{1}^{(i)}, \ldots, U_{N}^{(i)}\right)$ is the approximation of the solution at the $i$ th stage of the method, namely

$$
U_{j}^{(i)} \approx u\left(x_{j}, t_{n}+c_{i} \tau\right)
$$

in which $u$ is the solution of the original partial differential equation and $c_{i}=\sum_{j=1}^{m} a_{i j}$ (this latter condition is needed for the consistency of the method).

The area of mathematics in which strong stability preserving methods were defined was the theory of hyperbolic systems, in which the preservation of monoticity and positivity properties is an important requirement. The property which is usually guaranteed is that the scheme is total variation diminishing (or TVD for short), which in this case can be reformulated as the monotonicity condition

$$
\begin{equation*}
\left\|U^{n+1}\right\| \leq\left\|U^{n}\right\| \tag{1.35}
\end{equation*}
$$

The direct analysis of property (1.35) for Runge-Kutta methods (especially the ones written in the Butcher form) is usually hard. Instead of this, the usual process of the proof of such conditions consists of the following two steps:

- Prove that for the forward Euler method condition holds if

$$
\begin{equation*}
0 \leq \tau \leq \tau_{F E} \tag{1.36}
\end{equation*}
$$

- Rewrite the Runge-Kutta method in a way that the guarantee of property 1.35 can be easily proved when

$$
0 \leq \tau \leq \mathcal{C} \tau_{F E}
$$

in which $\mathcal{C}$ is usually referred to as the $S S P$ (or strong stability preserving) coefficient.

The latter point can be achieved by rewriting the Runge-Kutta method into the ShuOsher form

$$
\begin{align*}
U^{(i)} & =v_{i} U^{n-1}+\sum_{j=1}^{m}\left(\alpha_{i j} U^{(j)}+\tau \beta_{i j} \mathcal{P}\left(U^{(j)}\right)\right), \quad 1 \leq i \leq m+1  \tag{1.37}\\
U^{n} & =U^{(m+1)}
\end{align*}
$$

in which the coefficients $\alpha_{i j}$ and $\beta_{i j}$ can be written into $(m+1) \times(m+1)$ matrices and $v_{i}$ can be written as a vector with length $m+1$. Such methods were introduced by Shu as total-variation diminishing (TVD) discretizations [92, and by Shu and Osher in relation to high order spatial discretizations 93, 94].

From now on we consider numerical methods which have the following property.
Definition 1.6.1. A Runge-Kutta method (or any other numerical scheme) is said to be zero-well defined, if it gives a unique solution to the equation

$$
\left\{\begin{array}{l}
u^{\prime}(t)=0 \\
u(0)=u_{0}
\end{array}\right.
$$

The next theorem states that the Shu-Osher form is indeed a good choice for representation.

Theorem 1.6.1. 54, Theorem 2.1] Let us consider the initial-value problem (1.33), and assume that the operator $G(Q)$ satisfies the condition (1.35) applied to the forward Euler method, i.e.

$$
\|u+\tau \mathcal{P}(u)\| \leq\|u\| \quad \forall u
$$

for such a time step for which $0 \leq \tau \leq \tau_{F E}$ holds.
Now consider a zero-well defined Runge-Kutta method (written in the Shu-Osher form), and apply it to the initial-value problem (1.33). Then the numerical solution computed by this scheme satisfies the monotonicity condition (1.35) if

$$
0 \leq \tau \leq \mathcal{C}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \tau_{F E}
$$

in which $\mathcal{C}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is the SSP coefficient defined as

$$
\mathcal{C}(\boldsymbol{\alpha}, \boldsymbol{\beta})=\left\{\begin{array}{cl}
\min _{i, j} \frac{\alpha_{i j}}{\beta_{i j}}, & \text { if } \alpha_{i j}, \beta_{i j} \text { are non-negative } \\
0, & \text { otherwise }
\end{array}\right.
$$

One can prove that if we denote by $\mathcal{K}$ the matrix

$$
\mathcal{K}=\left[\begin{array}{cc}
\left(a_{i j}\right) & 0 \\
\boldsymbol{b}^{\top} & 0
\end{array}\right]
$$

(in which $a_{i j}$ and $\boldsymbol{b}$ are the coefficients of the Butcher form) and by $I$ the $(m+1)$ dimensional identity matrix, then a connection between the two representations can be established using the rule

$$
\mathcal{K}=(I-\boldsymbol{\alpha})^{-1} \boldsymbol{\beta}
$$

in which $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the matrices containing the coefficients of the Shu-Osher representation. It can also be shown that for zero-well defined methods the inverse $(I-\boldsymbol{\alpha})^{-1}$ exists.

It is evident that for a given Butcher form infinitely many Shu-Osher forms can be found. However, by Theorem 1.6 .1 we would like to have the representation with the largest value of SSP coefficient, thus obtaining the largest possible bound for our time step.

Let us assume that for the coefficients of the Shu-Osher representation, the relation

$$
p=\frac{\alpha_{i j}}{\beta_{i j}}, \quad \forall i, j
$$

holds, and from now on we use the notations for the matrices $\boldsymbol{\alpha}_{p}$ and $\boldsymbol{\beta}_{p}$ for which $\boldsymbol{\alpha}_{p}=p \boldsymbol{\beta}_{p}$. Then the coefficients of this special representation can be computed from the

Butcher coefficients using the rules

$$
\begin{aligned}
\boldsymbol{\beta}_{p} & =\mathcal{K}(I+p \mathcal{K})^{-1}, \\
\boldsymbol{\alpha}_{p} & =p \mathcal{K}(I+p \mathcal{K})^{-1}, \\
v_{p} & =(I+p \mathcal{K})^{-1} \boldsymbol{e},
\end{aligned}
$$

in which $\boldsymbol{e}$ is the all-one vector with length $m+1$. Using these, one can construct the canonical Shu-Osher form

$$
\begin{align*}
U^{(i)} & =v_{i} U^{n-1}+\sum_{j=1}^{m} \alpha_{i j}\left(U^{(j)}+\frac{\tau}{p} \mathcal{P}\left(U^{(j)}\right)\right), \quad 1 \leq i \leq m+1  \tag{1.38}\\
U^{n} & =U^{(m+1)}
\end{align*}
$$

in which $v_{i}$ and $\alpha_{i j}$ are the elements of $v_{p}$ and $\alpha_{p}$, respectively.
The Shu-Osher representation with the largest value of $p$ such that $(I+p \mathcal{K})^{-1}$ exists and $\boldsymbol{\alpha}_{p}, \boldsymbol{v}_{p}$ have non-negative components is called optimal and attains the SSP coefficient

$$
\mathcal{C}=\max \left\{p \geq 0 \mid \exists(I+p \mathcal{K})^{-1} \text { and } \boldsymbol{\alpha}_{p} \geq 0, \boldsymbol{v}_{p} \geq 0\right\}
$$

The interested reader may consult [52, 53, 54], as well as the monograph [55] and the references within, for a throughout review of SSP methods.

### 1.6.2 The application of SSP-RK methods

We would like to investigate time step restrictions such that the numerical solution obtained by applying method (1.38) to the problem (1.16) satisfies properties $D_{1}-D_{4}$. The following theorem provides the theoretical upper bound for the time step such that these properties are satisfied.

Theorem 1.6.2. Consider the numerical solution obtained by applying an explicit RungeKutta method (1.38) with SSP coefficient $\mathcal{C}>0$ to the semi-discrete problem 1.16 with non-negative initial data. Then property $D_{2}$ holds without any time step restrictions. Moreover, the properties $D_{1}, D_{3}$ and $D_{4}$ hold if the time step satisfies

$$
\begin{equation*}
\tau \leq \mathcal{C} \min \left\{\frac{1}{\widehat{T}+c}, \frac{1}{b}\right\} \tag{1.39}
\end{equation*}
$$

where $\widehat{T}$ is given by (1.27).
For the proof, see [104].
Note that the bound (1.39) is similar to the usual ones which are expected in the case of strong stability-preserving Runge-Kutta methods, since the bound for which the forward Euler method had properties $D_{1}-D_{4}$ was

$$
\tau \leq \widehat{\tau}:=\min \left\{\frac{1}{\widehat{T}+c}, \frac{1}{b}\right\}
$$

### 1.6.3 Numerical results

In this section we confirm the results proved in the previous sections by using several numerical experiments. Computational tests are defined in a bounded domain and thus the choice of boundary conditions is important. Because we have no diffusion in our problem, we consider homogeneous Dirichlet conditions and we assume that there is no susceptible population outside of our domain. This means that we are going to assign a zero value to any point which lies outside of the rectangular domain in which the problem is defined. In most cases the nodes of the cubatures rules 1.20 and 1.22 do not belong to the spatial grid. Special attention must be given to the corners and boundaries of the domain where cubature nodes, assigned to grid points near the boundary, lie outside of the domain. To be able to handle solution estimates at corners and at the boundary of the domain, we use ghost cells which are set to zero. This enables us to calculate the values corresponding to the cubature nodes lying outside of the domain.

For the numerical experiments we are choosing the following functions. Let $g_{1}(r)$ be a linearly decreasing function, which takes its maximum at $r=0$ and becomes zero at $r=\delta$, i.e.,

$$
g_{1}(r):=a(-r+\delta),
$$

where $a$ is similar to the parameter $a$ in (II). Also, we are going to use a non-constant symmetrical $g_{2}(\theta)$ function given by

$$
g_{2}(\theta):=\beta \sin (\theta+\alpha)+\beta .
$$

From now on, we are using the choices of $\alpha=0$ and $\beta=1$, in other words assuming a northern wind on the domain. In all numerical experiments - unless otherwise stated - we
use the parameter values $a=100, b=0.05, c=0.01$, and $\delta=0.05$, with 30 grid points in each direction and $6 \times 12$ cubature nodes. We also choose the tenth-stage, fourth-order SSP Runge-Kutta method (SSPRK104) for the time integration.

The initial conditions resemble the eruption of a wildfire, i.e., having infected cases located in a small area. For the infected species, we use a Gaussian distribution concentrated at the middle point $\left(\mathcal{L}_{1} / 2, \mathcal{L}_{2} / 2\right)$ of the domain $\Omega:=\left[0, \mathcal{L}_{1}\right] \times\left[0, \mathcal{L}_{2}\right]$, with standard deviation $\tilde{\sigma}=\min \left\{\mathcal{L}_{1}, \mathcal{L}_{2}\right\} / 10$. The spatial step sizes are $h_{1}=\mathcal{L}_{1} /\left(P_{1}-1\right)$ and $h_{2}=\mathcal{L}_{2} /\left(P_{2}-1\right)$, where $P_{1}$ and $P_{2}$ are the number of grid points in each direction. In all numerical tests we set $\mathcal{L}_{1}=\mathcal{L}_{2}=1$. We assume that the number of susceptibles is constant except at the middle of the domain, and there are no recovered species at the beginning. Therefore, for every $1 \leq k \leq P_{1}, 1 \leq l \leq P_{2}$ the initial conditions are given by

$$
\begin{aligned}
I_{k, \ell}^{0} & =\frac{1}{2 \pi \tilde{\sigma}^{2}} \exp \left(-\frac{1}{2}\left[\left(\frac{h_{1}(k-1)-\frac{\mathcal{L}_{1}}{2}}{\tilde{\sigma}}\right)^{2}+\left(\frac{h_{2}(l-1)-\frac{\mathcal{L}_{2}}{2}}{\tilde{\sigma}}\right)^{2}\right]\right) \\
S_{k, \ell}^{0} & =\frac{1}{2 \pi \tilde{\sigma}^{2}}-I_{k, \ell}^{0} \\
R_{k, \ell}^{0} & =0
\end{aligned}
$$

First we would like to study the behavior of our numerical solution. Figure 1.3 depicts the numerical solution at times $t=50$ and $t=500$. As we can see, the number of susceptibles is decreased, and the number of infected moves towards the boundaries, while forming a wave. Both densities $S$ and $I$ tend to zero, which confirms that the zero solution is indeed an asymptotically stable equilibrium for the first two equations of (1.3), as it was proved in Section 1.1.

### 1.6.3.1 Comparison of the step size bounds for the Euler method

As we saw in Section 1.5, the improved bound $\widehat{\tau}$ (see 1.29 ) is larger than the pessimistic bound $\widetilde{\tau}$ (see $(1.30)$ ), and thus closer to the best theoretically bound (1.25) that guarantees the preservation of properties $D_{1}-D_{4}$. We would like to determine how close the bound $\widehat{\tau}$ is to the adaptive step-size restriction, and compare it with the pessimistic bound $\widetilde{\tau}$. In Table 1.1 we have tested several different values of $a$ and $\delta$, for which both the bounds $\widehat{\tau}$ and $\widetilde{\tau}$ were computed. For comparison we calculated the minimum of the adaptive


Figure 1.3: The number of susceptibles $S$ (left), infected $I$ (middle) and recovered $R$ (right) at times $t=50$ (top panel) and $t=500$ (bottom panel). The Gauss-Legendre quadrature (1.22) has been used combined with the modified Akima ('makima') interpolation.
step bound (1.25), denoted by $\tau_{e}$. As we can see, varying the parameter $a$ and using the time step bound $\widehat{\tau}$ results in about $55 \%$ increase in efficiency and is much closer to the theoretical bound for which the properties $D_{1}-D_{4}$ hold. By varying the parameter $\delta$ instead of $a$, the time-step ratios remain similar and result in more than $70 \%$ difference between the improved bound (1.29) and the time step 1.30. From Table 1.1 we conclude that in the case of a small increase in the time step $\widehat{\tau}$, the forward Euler method continues to preserve the desired properties. However, for values of $\tau$ bigger than (1.39), there is

| $a$ | $\widetilde{\tau}$ | $\widetilde{\tau} / \tau_{e}$ | $\widehat{\tau}$ | $\widehat{\tau} / \tau_{e}$ | $\tau_{e}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 50 | 3.7458 | 0.4037 | 8.7682 | 0.9449 | 9.2792 |
| 100 | 1.9086 | 0.3923 | 4.5851 | 0.9424 | 4.8653 |
| 250 | 0.7723 | 0.3853 | 1.8859 | 0.9408 | 2.0046 |
| 500 | 0.3876 | 0.3857 | 0.9519 | 0.9470 | 1.0052 |
|  |  |  |  |  |  |
| $\delta$ | $\widetilde{\tau}$ | $\widetilde{\tau} / \tau_{e}$ | $\widehat{\tau}$ | $\widehat{\tau} / \tau_{e}$ | $\tau_{e}$ |
| 0.025 | 7.5188 | 0.3759 | 20.0 | 1.0 | 20.0 |
| 0.05 | 1.0060 | 0.2066 | 4.5802 | 0.9404 | 4.8703 |
| 0.075 | 0.3002 | 0.1959 | 1.4023 | 0.9151 | 1.5324 |
| 0.1 | 0.1269 | 0.2002 | 0.5964 | 0.9412 | 0.6337 |

Table 1.1: Step-size bounds $\widehat{\tau}$ and $\widetilde{\tau}$ (see (1.29) and (1.30) respectively), and their comparison with the adaptive bound $\tau_{e}$ (see 1.25) for the forward Euler method for different values of $a$ and $\delta$. The computation uses the Elhay-Kautsky cubature rule 1.22) combined with bilinear interpolation, and the final time is $t_{\mathrm{f}}=100$.
no guarantee that properties $D_{1}-D_{4}$ will be satisfied by a high-order time integration method.

### 1.6.3.2 Convergence of the method

Since we cannot approximate the exact solution accurately, we are going to compute the numerical errors for different methods by using a reference solution. To have a fair comparison the reference solution is computed by using the same parameters and method, but with either a large number of cubature nodes or a very small time step.

First, we observe how well the different cubatures behave. As seen in Section 1.4, using more nodes in cubature (1.22) results in smaller errors, and also faster convergence. Numerical experiments show that this is also the case for the system (1.16). The $L^{2}$-norm errors for the different cubature formulas and interpolations can be seen in Figure 1.4. It is clear that for a small number of cubature nodes there is no remarkable difference between the interpolations, but for more cubature nodes modified Akima and spline interpolation perform better. Bilinear interpolation results in similar errors for both cubatures 1.20 and (1.22). As it can be seen, mofified Akima and spline interpolation perform the same way for the Elhay-Kautsky cubature 1.20 and smaller errors are observed with spline interpolation and Gauss-Legendre cubature (1.22).

Equally important is the order of the different time integration methods. Table 1.2 shows that the forward Euler method behaves similarly when compared to the firstorder integral solution described in Section 1.3. Numerical experiments show that the


Figure 1.4: $L^{2}$-norm errors using cubatures formulas 1.20 and 1.22 with $n \times 2 n$ cubature nodes, $n \in\{3,4,6,9,12\}$ and different interpolations. The final time is $t_{\mathrm{f}}=50$ and the reference solution for each cubature rule and interpolation is computed by using $17 \times 34$ cubature nodes.
higher order schemes work as expected, namely that by using enough cubature nodes and grid points, a reasonably small error can be achieved with the desired accuracy order. Table 1.3 shows the convergence rates for second-, third- and fourth-order SSP Runge-Kutta methods when the Gauss-Legendre quadrature rule 1.22 is used with spline interpolation. The numerical solution is computed at time $t_{f}=50$ using 30 grid points and $6 \times 12$ cubature nodes. We start with a reasonable time step 4.7 , which is slightly below the minimum of the adaptive bound 1.25 when forward Euler method is used, and then successively divide by 2 . For the reference solution we use a time step that is the half of the smallest time step in our computations. It is evident that using higher order methods is better than solving the integral equation (1.12) numerically. Moreover the fourth-order SSP Runge-Kutta method (SSPRK104) attends a six times larger time step than lower order methods since it has an SSP coefficient $\mathcal{C}=6$.

### 1.7 Discretization in time: Operator splitting techniques

The main novelty of this section is (besides the traditional time discretization) that we use another time discretization-like method: operator splitting. As one can see, the

| $\tau$ | FE |  | IM |  |
| :---: | :---: | :---: | :---: | :---: |
| 1.0000 | $3.58 \times 10^{-1}$ |  | $8.17 \times 10^{-1}$ |  |
| 0.5000 | $1.82 \times 10^{-1}$ | 0.98 | $4.75 \times 10^{-1}$ | 0.78 |
| 0.2500 | $8.92 \times 10^{-2}$ | 1.03 | $2.53 \times 10^{-1}$ | 0.91 |
| 0.1250 | $4.19 \times 10^{-2}$ | 1.09 | $1.24 \times 10^{-1}$ | 1.02 |
| 0.0625 | $1.80 \times 10^{-2}$ | 1.22 | $5.48 \times 10^{-2}$ | 1.18 |

Table 1.2: $L^{2}$-norm errors and convergence rates of forward Euler method (FE) and the method (1.14), denoted by "IM". The solution is computed at time $t_{\mathrm{f}}=50$ with the Gauss-Legendre quadrature rule (1.22) combined with spline interpolation.

| SSPRK22 |  | SSPRK33 |  | SSPRK104 |  |  |
| :---: | :--- | :--- | :--- | :--- | :--- | :--- |
| 4.7000 | $3.35 \times 10^{-1}$ | $6.22 \times 10^{-2}$ |  | $8.99 \times 10^{-4}$ |  |  |
| 2.3500 | $1.07 \times 10^{-1}$ | 1.65 | $1.05 \times 10^{-2}$ | 2.57 | $6.46 \times 10^{-5}$ | 3.80 |
| 1.1750 | $3.03 \times 10^{-2}$ | 1.82 | $1.53 \times 10^{-3}$ | 2.78 | $4.31 \times 10^{-6}$ | 3.91 |
| 0.5875 | $8.01 \times 10^{-3}$ | 1.92 | $2.07 \times 10^{-4}$ | 2.89 | $2.78 \times 10^{-7}$ | 3.95 |
| 0.2938 | $1.97 \times 10^{-3}$ | 2.02 | $2.65 \times 10^{-5}$ | 2.96 | $1.76 \times 10^{-8}$ | 3.98 |
| 0.1469 | $4.01 \times 10^{-4}$ | 2.30 | $3.00 \times 10^{-6}$ | 3.14 | $1.04 \times 10^{-9}$ | 4.08 |

Table 1.3: $L^{2}$-norm errors and convergence rates of high-order integration methods. The solution is computed at time $t_{\mathrm{f}}=50$ with the Gauss-Legendre quadrature rule (1.22) combined with spline interpolation.
right-hand side of problem (1.3) can be written as a sum of two terms: one containing the integral and one with the remaining terms. The idea of operator splitting is to "split" the problem into two sub-problems with the corresponding terms alone, and solve them separately by using an appropriate initial condition to link their solution together. In the present section we will introduce and study the sequential, the sequential weighted, and the Strang splitting schemes.

As already mentioned, it is natural to split the space-discretized SIR model (1.16) into the sub-problems with and without the integral term $T_{k, \ell}$ specifying the space-dependency of the infection process:

$$
\left\{\begin{align*}
\partial_{t} S_{k, \ell}^{[1]}(t) & =-c S_{k, \ell}^{[1]}(t),  \tag{Sub.1}\\
\partial_{t} I_{k, \ell}^{[1]}(t) & =-b I_{k, \ell}^{[1]}(t), \\
\partial_{t} R_{k, \ell}^{[1]}(t) & =b I_{k, \ell}^{[1]}(t)+c S_{k, \ell}^{[1]}(t)
\end{align*}\right.
$$

and

$$
\left\{\begin{align*}
\partial_{t} S_{k, \ell}^{[2]}(t) & =-S_{k, \ell}^{[2]}(t) T_{k, \ell}^{[2]}\left(t, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right)  \tag{Sub.2}\\
\partial_{t} I_{k, \ell}^{[2]}(t) & =S_{k, \ell}^{[2]}(t) T_{k, \ell}^{[2]}\left(t, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right) \\
\partial_{t} R_{k, \ell}^{[2]}(t) & =0
\end{align*}\right.
$$

for all $t \geq 0$ and $(x, y) \in \Omega$. The link between the sub-problems is the initial condition, as will be shown in the next sections.

For the later use we remark that sub-problem Sub.1) can be solved exactly:

$$
\left\{\begin{align*}
S_{k, \ell}^{[1]}(t+\Delta t) & =\mathrm{e}^{-c \Delta t} S_{k, \ell}^{[1]}(t),  \tag{1.40}\\
I_{k, \ell}^{[1]}(t+\Delta t) & =\mathrm{e}^{-b \Delta t}\left[l_{k, \ell}^{[1]}(t),\right. \\
R_{k, \ell}^{[1]}(t+\Delta t) & =R_{k, \ell}^{[1]}(t)+\left(1-\mathrm{e}^{-c \Delta t}\right) S_{k, \ell}^{[1]}(t)+\left(1-\mathrm{e}^{-b \Delta t}\right) I_{k, \ell}^{[1]}(t)
\end{align*}\right.
$$

for all $t \geq 0$ and $(x, y) \in \Omega$, where $\Delta t \geq 0$ is an arbitrary time difference.
On the other hand, sub-problem (Sub.2) cannot be solved exactly. Its approximate solution can be obtained by another time discretization method. For instance, the use of the first-order explicit Euler method with time step $\tau>0$ leads to

$$
\left\{\begin{align*}
S_{k, \ell}^{[2]}((n+1) \tau) & =S_{k, \ell}^{[2]}(n \tau)-\tau S_{k, \ell}^{[2]}(n \tau) T_{k, \ell}^{[2]}\left(n \tau, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right)  \tag{1.41}\\
I_{k, \ell}^{[2]}((n+1) \tau) & =I_{k, \ell}^{[2]}(n \tau)+\tau S_{k, \ell}^{[2]}(n \tau) T_{k, \ell}^{[2]}\left(n \tau, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right), \\
R_{k, \ell}^{[2]}((n+1) \tau) & =R_{k, \ell}^{[2]}(n \tau)
\end{align*}\right.
$$

for all $n \in \mathbb{N}$ with $X^{[2]}(0, x, y)=X_{0}(x, y)$ for each $X \in\{S, I, R\}$. We note that we take $0 \in \mathbb{N}$.

The use of the second-order Heun's method in Shu-Osher form (which preserves the strong stability, see [52]) with time step $\tau>0$ results in the following steps:

$$
\begin{align*}
& \left\{\begin{array}{l}
\widehat{S}_{k, \ell}^{[2]}((n+1) \tau)=S_{k, \ell}^{[2]}(n \tau)-\tau S_{k, \ell}^{[2]}(n \tau) T_{k, \ell}^{[2]}\left(n \tau, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right), \\
\widehat{I}_{k, \ell}^{[2]}((n+1) \tau)=I_{k, \ell}^{[2]}(n \tau)+\tau S_{k, \ell}^{[2]}(n \tau) T_{k, \ell}^{[2]}\left(n \tau, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right), \\
\widehat{R}_{k, \ell}^{[2]}((n+1) \tau)=R_{k, \ell}^{[2]}(n \tau),
\end{array}\right.  \tag{1.42}\\
& \left\{\begin{aligned}
S_{k, \ell}^{[2]}((n+1) \tau) & =\frac{1}{2} S_{k, \ell}^{[2]}(n \tau) \\
& \left.+\frac{1}{2}\left(\widehat{S}_{k, \ell}^{[2]}((n+1) \tau)-\tau \widehat{S}_{k, \ell}^{[2]}((n+1) \tau) \widehat{T}_{k, \ell}^{[2]}(n+1) n \tau, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right)\right), \\
I_{k, \ell}^{[2]}((n+1) \tau) & =\frac{1}{2} I_{k, \ell}^{[2]}(n \tau) \\
& +\frac{1}{2}\left(\widehat{I}_{k, \ell}^{[2]}((n+1) \tau)+\tau \widehat{S}_{k, \ell}^{[2]}((n+1) \tau) \widehat{T}_{k, \ell}^{[2]}\left((n+1) \tau, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right)\right), \\
R_{k, \ell}^{[2]}((n+1) \tau) & =\frac{1}{2} R_{k, \ell}^{[2]}(n \tau)+\frac{1}{2} \widehat{R}_{k, \ell}^{[2]}((n+1) \tau)=R_{k, \ell}^{[2]}(n \tau),
\end{aligned}\right. \tag{1.43}
\end{align*}
$$

where $\widehat{T}_{k, \ell}$ is calculated similarly as $T_{k, \ell}$, but with the use of $\widehat{I}_{k, \ell}$ instead of $I_{k, \ell}$. We do not plug formulae (1.42) into (1.43), because the method will be more suitable for analysis
in its present form.
We note that the use of an additional time discretization inside one time step might also lead to a positivity preserving method. This technique may be applied only for one sub-problem. Then the time step could be chosen independently of the constraints but related to the accuracy of the scheme. For more details on such kind of adaptive time stepping we refer to [40].

### 1.7.1 Technical tools

Before the derivation and analysis of the methods, we collect some notations and technical tools we will use later on.

Notation 1.7.1. 1. Let $\mathcal{M}: \mathbb{R}^{P_{1} P_{2} \times P_{1} P_{2}} \rightarrow \mathbb{R}^{P_{1} P_{2} \times P_{1} P_{2}}$ denote the bounded linear operator (represented by a matrix in applications) that maps $I^{n}$ to $T^{n}$ by the rule $T^{n}=\mathcal{M}\left(I^{n}\right)$. Furthermore, let

$$
M:=\|\mathcal{M}\|_{\infty} \cdot\left\|S^{0}+I^{0}+R^{0}\right\|_{\infty}
$$

in which $\|.\|_{\infty}$ means the maximum matrix norm taken element-wise. Note that from now on we assume that the condition

$$
S^{0}(x, y)+I^{0}(x, y)+R^{0}(x, y) \neq 0 \quad \forall(x, y) \in \Omega
$$

holds, which implies $M>0$.
2. Let $W_{-1}:[-1 / \mathrm{e}, 0) \rightarrow(-\infty,-1]$ and $W_{0}:(-1 / \mathrm{e},+\infty) \rightarrow(-1,+\infty)$ denote the two branches of the Lambert- $W$ function, that is, the inverse of the map $x \mapsto x \mathrm{e}^{x}$.
3. For arbitrary $p, q>0$, we define the set

$$
\mathbb{T}_{p, q}:=\left[0,-\frac{1}{p} W_{0}\left(-\frac{p}{q}\right)\right] \cup\left[-\frac{1}{p} W_{-1}\left(-\frac{p}{q}\right),+\infty\right) \subset \mathbb{R}
$$

Furthermore, we define

$$
\mathbb{T}_{0, q}:=\left[0, \frac{1}{q}\right) \subset \mathbb{R}
$$

The latter notation makes sense because of the following consideration.

Lemma 1.7.1. With Notation 1.7.1, the limit $-\frac{1}{p} W_{0}\left(-\frac{p}{q}\right) \xrightarrow{p \rightarrow 0} \frac{1}{q}$ holds for arbitrary $q>0$.

Proof. It suffices to show that $W_{0}(x) / x \xrightarrow{x \rightarrow 0} 1$ for $x=-p / q<0$. The L'Hospital rule, the derivative of the inverse function, and the identity $W_{0}(0)=0$ imply that

$$
\lim _{x \rightarrow 0} \frac{W_{0}(x)}{x}=\lim _{x \rightarrow 0} W_{0}^{\prime}(x)=\lim _{x \rightarrow 0} \frac{1}{\mathrm{e}^{W_{0}(x)}+W_{0}(x) \mathrm{e}^{W_{0}(x)}}=1
$$

Remark 1.7.1. Since we will use it several times throughout the paper, we analyse the solution $x<0$ to equation

$$
\begin{equation*}
x \mathrm{e}^{x}=\rho \tag{1.44}
\end{equation*}
$$

for some parameter $\rho<0$.

1. For $\rho<-1 / \mathrm{e}$, there is no solution to equation (1.44).
2. For $\rho=-1 / \mathrm{e}$, there is one solution: $x_{1}=-1$.
3. For $\rho>-1 / \mathrm{e}$, there are two solutions: $x_{-1}=W_{-1}(\mu)$ and $x_{0}=W_{0}(\mu)$.

We also know that $x_{-1} \leq x_{1}=-1<x_{0}$. Hence, for the inequality

$$
\begin{equation*}
x \mathrm{e}^{x} \geq \rho \tag{1.45}
\end{equation*}
$$

we have the following cases.

1. For $\rho<-1 /$ e, the inequality 1.45 holds for every $x<0$.
2. For $\rho=-1 / \mathrm{e}$, the inequality 1.45) holds for every $x<0$ (we have $x \mathrm{e}^{x}=\rho$ for $x=-1$ ).
3. For $\rho>-1 / \mathrm{e}$, we have: $x<x_{-1}=W_{-1}(\mu)$ or $x>x_{0}=W_{-1}(\mu)$.

The graph of function $x \mapsto x \mathrm{e}^{x}$ is depicted in Figure 1.5.
In the next sections we will present the condition on the time step $\tau$ under which the qualitative properties $D_{1}-D_{4}$ hold for the various operator splitting schemes. We are especially interested in the cases when the application of operator splitting leads to less


Figure 1.5: Graph of function $x \mapsto x \mathrm{e}^{x}$. The horizontal lines indicate the $\rho$-values -0.25 and $-1 / \mathrm{e}$.
severe condition than the one obtained without splitting. As it was mentioned in Section 1.5, in [103] the authors applied the explicit Euler method without taking into account the vaccination $(c=0)$. They found that property $D_{2}$ was automatically satisfied, and properties $D_{1}, D_{3}$ and $D_{4}$ held true for time steps $\tau$ satisfying

$$
\tau \leq \min \left\{\frac{1}{M}, \frac{1}{b}\right\}
$$

The case $c>0$ was studied in [104, and resulted in a similar bound, namely

$$
\begin{equation*}
\tau \leq \min \left\{\frac{1}{M+c}, \frac{1}{b}\right\} \tag{1.46}
\end{equation*}
$$

From now on, the upper bound (1.46) will be considered as a reference value, and we will study the conditions under which the application of operator splitting procedures leads to a higher one.

### 1.7.2 Sequential splitting

Operator splitting is based on the idea of simplifying the problem by splitting it into two or more sub-problems which are easier to solve or treat numerically. Since the subproblems need to be solved separately, we should derive a way to connect their solutions. Depending on these rules, we distinguish several splitting methods. The most basic one is the sequential splitting (initiated first in [10]) when the sub-problems are solved one after
the other on a time interval of length $\tau>0$, always taking the solution of the previous sub-problem as initial condition for the actual one. As we will see, the properties of the sequential splitting depend on the order of the sub-problems, therefore, we will treat the two different cases arising from the different orders separately.

Another splitting procedure is derived when the solutions of the two types of sequential splittings are weighted by a parameter $\Theta \in(0,1)$. This kind of method is called weighted sequential splitting (see [31]) and will be discussed in Section 1.7.3. The third operator splitting to be discussed in Section 1.7 .4 is the Strang splitting (derived in [96] and [77]) solving three problems in a single time step: one with the first sub-problem over a time interval of length $\tau / 2$, then with the second sub-problem on an interval of length $\tau$, and finally with the first sub-problem again on a $\tau / 2$ interval.

In what follows we analyse the splitting procedures in the light of whether they preserve the qualitative properties $D_{1}-D_{4}$.

### 1.7.2.1 Sequential splitting $1-2$

First we treat the sequential splitting in the case when the sub-problems are taken in the order (Sub.1)-(Sub.2). Then the application of the sequential splitting means that in a single time step we first solve sub-problem (Sub.1) whose solution 1.40) serves as the initial condition to sub-problem Sub.2):

$$
\begin{align*}
& X^{[1]}(n \tau, x, y)=X^{[2]}(n \tau, x, y),  \tag{1.47}\\
& X^{[2]}(n \tau, x, y)=X^{[1]}((n+1) \tau, x, y)
\end{align*}
$$

for all $n \in \mathbb{N}$ with $\tau>0$, where $X^{[2]}(0, x, y)=X_{0}(x, y)$ is the original initial values of the solution at $t=0$ for each $X \in\{S, I, R\}$. After discretizing sub-problem Sub.2 by the explicit Euler method and discretizing sub-problems (1.40) and (1.41) in time, we get the following two sub-problems:

$$
\left\{\begin{align*}
S_{k, \ell}^{[1], n+1} & =\mathrm{e}^{-c \tau} S_{k \ell}^{[1], n},  \tag{1.48}\\
I_{k, \ell}^{[1], n+1} & =\mathrm{e}^{-b \tau} I_{k, \ell}^{[1], n} \\
R_{k, \ell}^{[1], n+1} & =R_{k, \ell}^{[1], n}+\left(1-\mathrm{e}^{-c \tau}\right) S_{k, \ell}^{[1], n}+\left(1-\mathrm{e}^{-b \tau}\right) I_{k, \ell}^{[1], n}
\end{align*}\right.
$$

and

$$
\left\{\begin{align*}
S_{k, \ell}^{[2], n+1} & =S_{k, \ell}^{[2], n}-\tau S_{k, \ell}^{[2], n} T_{k, \ell}^{[2], n},  \tag{1.49}\\
I_{k, \ell}^{[2], n+1} & =I_{k, \ell}^{[2], n}+\tau S_{k, \ell}^{[2], n} T_{k, \ell}^{[2], n}, \\
R_{k, \ell}^{[2], n+1} & =R_{k, \ell}^{[2], n} .
\end{align*}\right.
$$

By taking into the initial conditions (1.47), the sub-problems have the following form for all $n \in \mathbb{N}$ and given $S_{k, \ell}^{n}, I_{k, \ell}^{n}, R_{k, \ell}^{n}$ :

$$
\begin{align*}
& \left\{\begin{array}{l}
S_{k, \ell}^{[1], n+1}=\mathrm{e}^{-c \tau} S_{k, \ell}^{n}, \\
I_{k, \ell}^{[1], n+1}=\mathrm{e}^{-b \tau} I_{k, \ell}^{n}, \\
R_{k, \ell}^{[1], n+1}=R_{k, \ell}^{n}+\left(1-\mathrm{e}^{-c \tau}\right) S_{k, \ell}^{n}+\left(1-\mathrm{e}^{-b \tau}\right) I_{k, \ell}^{n}, \\
\left\{\begin{array}{l}
S_{k, \ell}^{n+1}=S_{k, \ell}^{[1], n+1}-\tau S_{k, \ell}^{[1], n+1} T_{k, \ell}^{[1], n+1}, \\
I_{k, \ell}^{n+1}=I_{k, \ell}^{[1], n+1}+\tau S_{k, \ell}^{[1], n+1} T_{k, \ell}^{[1], n+1}, \\
R_{k, \ell}^{n+1}=R_{k, \ell}^{[1], n+1} .
\end{array}\right.
\end{array} .\right. \tag{1.50}
\end{align*}
$$

Notation 1.7.1.(1) and the linearity of operator $\mathcal{M}$ imply the following relation:

$$
\begin{equation*}
T_{k, \ell}^{[1], n+1}=\mathcal{M}\left(I_{k, \ell}^{[2], n}\right)=\mathcal{M}\left(I_{k, \ell}^{[1], n+1}\right)=\mathcal{M}\left(\mathrm{e}^{-b \tau} I_{k, \ell}^{n}\right)=\mathrm{e}^{-b \tau} \mathcal{M}\left(I_{k, \ell}^{n}\right)=\mathrm{e}^{-b \tau} T_{k, \ell}^{n} . \tag{1.52}
\end{equation*}
$$

By combining the sub-problems (1.50) (1.51), and the relation (1.52), we arrive at the numerical scheme

$$
\left\{\begin{array}{l}
S_{k, \ell}^{n+1}=\mathrm{e}^{-c \tau} S_{k, \ell}^{n}\left(1-\tau \mathrm{e}^{-b \tau} T_{k, \ell}^{n}\right),  \tag{1.53}\\
I_{k, \ell}^{n+1}=\mathrm{e}^{-b \tau}\left(I_{k, \ell}^{n}+\tau \mathrm{e}^{-c \tau} S_{k, \ell}^{n} T_{k, \ell}^{n}\right), \\
R_{k, \ell}^{n+1}=R_{k, \ell}^{n}+\left(1-\mathrm{e}^{-c \tau}\right) S_{k, \ell}^{n}+\left(1-\mathrm{e}^{-b \tau}\right) I_{k, \ell}^{n} .
\end{array}\right.
$$

In what follows we show the connection between properties $D_{1}-D_{4}$, and investigate the conditions under which they are fulfilled.

Proposition 1.7.1. We have the following assertions.

1. Property $D_{2}$ holds for the numerical method (1.53) without any restriction.
2. Properties $D_{3}$ and $D_{4}$ are consequences of property $D_{1}$.

Proof. (1) Property $D_{2}$ follows by adding up the equations of system (1.53).
(2) Since $T_{k, \ell}^{n} \geq 0$ holds if $I_{k, \ell}^{n} \geq 0$, and $\mathrm{e}^{-b \tau}>0$ in the first and the third equations of system (1.53), we get that properties $D_{3}$ and $D_{4}$ also hold.
This concludes the proof.
Due to Proposition 1.7.1, the monotonicity properties $D_{3}$ and $D_{4}$ follow from the non-negativity property $D_{1}$. Thus, we do not need to treat them separately. Hence, as a next step we study the conditions under which the non-negativity property $D_{1}$ holds.

Proposition 1.7.2. With Notation 1.7.1, we have the following assertions.

1. For $M<b e$, the non-negativity property $D_{1}$ is satisfied for all values of time step $\tau>0$.
2. For $M \geq b e$, the non-negativity property $D_{1}$ holds if $\tau \in \mathbb{T}_{b, M}$.

For the proof, see [33].
Interestingly, the condition $\tau \in \mathbb{T}_{b, M}$ in Proposition 1.7 .2 means that there is a "forbidden interval"

$$
\left(-\frac{1}{b} W_{0}\left(-\frac{b}{M}\right),-\frac{1}{b} W_{-1}\left(-\frac{b}{M}\right)\right) \subset \mathbb{R}
$$

where $\tau$ leads to negative $S, I, R$ values. It is worth mentioning, however, that Proposition 1.7 .2 gives a necessary condition only, so the forbidden interval can be shorter in real applications. The correspondence between the "exact" and the necessary bounds will be investigated in Section 1.7.5.

It is important to compare the bounds obtained for the time step in Proposition 1.7.2 with the similar result obtained for a numerical method without using operator splitting, cf. bound (1.46).

Proposition 1.7.3. With Notation 1.7.1, we have the following assertions.

1. The estimate $-\frac{1}{b} W_{0}\left(-\frac{b}{M}\right)>\frac{1}{M+c}$ holds for all $M, b, c>0$ with $M>b e$.
2. For an arbitrary $M>0$, we have the limit $-\frac{1}{b} W_{0}\left(-\frac{b}{M}\right) \xrightarrow{b \rightarrow 0} \frac{1}{M}$.

Proof. 1. The relation $W_{0}(y)<y$ for all $y<0$, the strictly increasing property of $W_{0}$ and the assumption $M>$ be imply the assertion.
2. Follows from Lemma 1.7 .1 with $p=b$ and $q=M$.

Proposition 1.7 .3 means that in the case $M>$ be our method (1.53) gives a larger upper bound for the time step $\tau$ as the application of explicit Euler method without operator splitting. Namely, in this case $M+c>M>b e>b$ holds, which leads to $\min \{1 / b, 1 /(M+c)\}=1 /(M+c)$. Moreover, for the case $M \leq b e$, our method 1.53 ) satisfies the properties $D_{1}-D_{4}$ without any restriction on the time step $\tau$. Hence, method (1.53) is more convenient to use than the method proposed in Section 1.5 (and in [103]).

We note here, that although Proposition 1.7 .2 allows large values for the time step, the use of these is not advised, since it leads to considerable higher error in the numerical solution.

Remark 1.7.2. Another possible way to perform the time step analysis is to check the nonnegativity preservation for each sub-problem separately, and then take the most severe constraint on the time step. In case of the sequential splitting 1-2 (1.53) however, we obtain a weaker result than the one presented in Proposition 1.7.2, namely, $\tau \leq 1 / M$. This can be seen from the following consideration. Sub-problem (1.50) preserves the nonnegativity for all $\tau>0$, while sub-problem (1.56) introduces the constraint $\tau \leq 1 / M$. This bound is always smaller than the one obtained in Proposition 1.7.2, which can be seen from the proof of Proposition 1.7 .3 (1). Thus, the point in analysing the combined method (1.53) is that it might lead to sharper constrains on the time step, as it does in this case.

### 1.7.2.2 Sequential splitting $2-1$

We study now the sequential splitting with the other order of the sub-problems. In a single time step we first solve (Sub.2) and then (Sub.1) with the initial conditions

$$
\begin{align*}
& X^{[2]}(n \tau, x, y)=X^{[1]}(n \tau, x, y),  \tag{1.54}\\
& X^{[1]}(n \tau, x, y)=X^{[2]}((n+1) \tau, x, y)
\end{align*}
$$

for all $n \in \mathbb{N}$ and $(x, y) \in \Omega$ with $X^{[1]}(0, x, y)=X_{0}(x, y)$ for each $X \in\{S, I, R\}$. Thus, we consider first the space discretized sub-problem (1.49) and then (1.48). Then the
numerical method takes the form

$$
\begin{align*}
& \left\{\begin{array}{l}
S_{k, \ell}^{[1], n+1}=S_{k, \ell}^{n}-\tau S_{k, \ell}^{n} T_{k, \ell}^{n}, \\
I_{k, \ell}^{[1], n+1}=I_{k, \ell}^{n}+\tau S_{k, \ell}^{n} T_{k, \ell}^{n}, \\
R_{k, \ell}^{[1], n+1}=R_{k, \ell}^{n},
\end{array}\right.  \tag{1.55}\\
& \left\{\begin{array}{l}
S_{k, \ell}^{n+1}=\mathrm{e}^{-c \tau} S_{k, \ell}^{[1], n+1}, \\
I_{k, \ell}^{n+1}=\mathrm{e}^{-b \tau} I_{k, \ell}^{[1], n+1}, \\
R_{k, \ell}^{n+1}=R_{k, \ell}^{[1], n+1}+\left(1-\mathrm{e}^{-c \tau}\right) S_{k, \ell}^{[1], n+1}+\left(1-\mathrm{e}^{-b \tau}\right) I_{k, \ell}^{[1], n+1}
\end{array}\right. \tag{1.56}
\end{align*}
$$

for all $n \in \mathbb{N}$ and $\left(x_{k}, y_{\ell}\right) \in \mathcal{G}$. Combination of sub-problems (1.55) and (1.56) yields the method

$$
\left\{\begin{array}{l}
S_{k, \ell}^{n+1}=\mathrm{e}^{-c \tau} S_{k, \ell}^{n}\left(1-\tau T_{k, \ell}^{n}\right)  \tag{1.57}\\
I_{k, \ell}^{n+1}=\mathrm{e}^{-b \tau}\left(I_{k, \ell}^{n}+\tau S_{k, \ell}^{n} T_{k, \ell}^{n}\right) \\
R_{k, \ell}^{n+1}=R_{k, \ell}^{n}+\left(1-\mathrm{e}^{-c \tau}\right) S_{k, \ell}^{n}\left(1-\tau T_{k, \ell}^{n}\right)+\left(1-\mathrm{e}^{-b \tau}\right)\left(I_{k, \ell}^{n}+\tau S_{k, \ell}^{n} T_{k, \ell}^{n}\right)
\end{array}\right.
$$

We can state the same result as before.
Proposition 1.7.4. Proposition 1.7.1 holds for the method (1.57).
Proof. First, we add up the equations in 1.57 to obtain property $D_{2}$. To prove the next assertion, we consider an arbitrary step again. Property $D_{1}$ implies that $T_{k, \ell}^{n}$ is non-negative. This and $\mathrm{e}^{-c \tau}<1$ imply that $S_{k, \ell}^{n+1} \leq S_{k, \ell}^{n}$. Moreover, the non-negativity of $S_{k, \ell}^{n+1}$ implies that $1-\tau T_{k, \ell}^{n} \geq 0$, therefore, $R_{k, \ell}^{n+1} \geq R_{k, \ell}^{n}$ holds as well.

According to Proposition 1.7 .4 it suffices to show the non-negativity property $D_{1}$ to obtain the monotonicity properties $D_{3}$ and $D_{4}$.

Proposition 1.7.5. The non-negativity property $D_{1}$ holds true for the method (1.57) if the time step $\tau$ satisfies the condition

$$
\begin{equation*}
\tau \leq \frac{1}{M} \tag{1.58}
\end{equation*}
$$

where $M$ is defined in Notation 1.7.1.
Proof. Since the initial values are non-negative, we assume that $X_{k, \ell}^{n} \geq 0$ and show that $X_{k, \ell}^{n+1} \geq 0$ for all $n \in \mathbb{N},\left(x_{k}, y_{\ell}\right) \in \mathcal{G}$, and $X \in\{S, I, R\}$. Since the assumption $\tau \leq 1 / M$
implies $1-\tau T_{k, \ell}^{n} \geq 0$, the non-negativity of $S_{k, \ell}^{n+1}$ is fulfilled. Furthermore, since all additive terms in the second and third equations of (1.57) are non-negative, we have $I_{k, \ell}^{n+1} \geq 0$ as well as $R_{k, \ell}^{n+1} \geq 0$.

Hence, for $M+c>b$ we get a better bound for the time step $\tau$ than for the explicit Euler method without splitting, cf. (1.46). If $M+c<b$, it might happen that the bound of the non-split method is better. Since the application of operator splitting usually needs more CPU time than the explicit Euler method itself, it is not advised to use method (1.57) but the first one (1.53).

We note that in case of the sequential splitting 2-1 we obtain the same bound (1.58) on the time step $\tau$ both when analysing the combined method (1.57) or the separate sub-problems 1.55 and 1.56 .

### 1.7.3 Weighted sequential splitting

Especially on parallel computers, it is a good idea to combine the solutions obtained by using the sequential splittings $1-2$ and $2-1$ with some $\Theta \in[0,1]$ parameter as follows:

$$
X=\Theta \cdot X_{\underline{1.53)}}+(1-\Theta) \cdot X_{\underline{1.57}}
$$

where $X_{(1.53)}$ and $X_{(1.57)}$ denote the approximate solutions obtained by numerical methods (1.53) and 1.57), respectively, for each $X \in\{S, I, R\}$. We note that the choice $\Theta=0$ results in the method (1.57), while $\Theta=1$ gives (1.53). In this way we get the following numerical method:

$$
\left\{\begin{array}{l}
S_{k, \ell}^{n+1}=\mathrm{e}^{-c \tau} S_{k, \ell}^{n}\left(1-\tau\left(\Theta \mathrm{e}^{-b \tau}+(1-\Theta)\right) T_{k, \ell}^{n}\right)  \tag{1.59}\\
I_{k, \ell}^{n+1}=\mathrm{e}^{-b \tau}\left(I_{k, \ell}^{n}+\tau\left(\Theta \mathrm{e}^{-c \tau}+(1-\Theta)\right) S_{k, \ell}^{n} T_{k, \ell}^{n}\right) \\
R_{k, \ell}^{n+1}=R_{k, \ell}^{n}+\left(1-\mathrm{e}^{-c \tau}\right) S_{k, \ell}^{n}\left(1-(1-\Theta) \tau T_{k, \ell}^{n}\right)+\left(1-\mathrm{e}^{-b \tau}\right)\left(I_{k, \ell}^{n}+(1-\Theta) \tau S_{k, \ell}^{n} T_{k, \ell}^{n}\right)
\end{array}\right.
$$

As before, we investigate the validity of properties $D_{2}{ }^{-} D_{4}$.
Proposition 1.7.6. Proposition 1.7.1 is valid for method (1.59).
Proof. The conservation of the size of the population is obtained again by adding up the equations in 1.59. Since $\Theta, \mathrm{e}^{-b \tau}, \mathrm{e}^{-c \tau}>0$ and $T_{k, \ell}^{n} \geq 0$ in the first equation of 1.59),
we have $S_{k, \ell}^{n+1} \leq S_{k, \ell}^{n}$. Due to property $D_{1}$, all terms in the third equation of (1.59) are non-negative, therefore, $R_{k, \ell}^{n+1} \geq R_{k, \ell}^{n}$ holds true.

In order to study the non-negativity preservation $D_{1}$, we need the following notation.
Notation 1.7.2. For the parameter $\Theta \in[0,1]$, we define

$$
\Theta^{*}:=\frac{\mathrm{e}^{2}}{\mathrm{e}^{2}+1} \approx 0.8808 .
$$

It will turn out that we get remarkably different bounds for $\Theta$ being under or above $\Theta^{*}$.

Notation 1.7.3. 1. For $\Theta \in[0,1]$ and $b>0$, we define the function $V_{\Theta, b}: \mathbb{R}^{+} \rightarrow(0,+\infty)$ as

$$
V_{\Theta, b}(\tau)=\tau\left(1-\Theta\left(1-\mathrm{e}^{-b \tau}\right)\right) .
$$

2. We introduce the values $0<\tau_{0}<\tau_{-1}$ as

$$
\begin{aligned}
\tau_{-1} & :=\frac{1}{b}\left(1-W_{-1}\left(\frac{\mathrm{e}(\Theta-1)}{\Theta}\right)\right) \quad \text { for } \quad \Theta \in\left[\Theta^{*}, 1\right), \\
\tau_{0} & :=\frac{1}{b}\left(1-W_{0}\left(\frac{\mathrm{e}(\Theta-1)}{\Theta}\right)\right) \quad \text { for } \quad \Theta \in\left[\Theta^{*}, 1\right] .
\end{aligned}
$$

in Figure 1.6 the graph of function $V_{\Theta, b}$ is shown for $\Theta=0.95$ and $b=0.1$. In order to illustrate its dependence on $\Theta$, we present the graph of function $V_{\Theta, b}$ for various values of $\Theta$ and $b=0.1$ in Figure 1.7 .


Figure 1.6: Graph of function $V_{\Theta, b}(\tau)=\tau\left(1-\Theta\left(1-\mathrm{e}^{-b \tau}\right)\right)$ for $\Theta=0.95$ and $b=0.1$. The horizontal line indicates the value 3.5.


Figure 1.7: Graph of function $V_{\Theta, b}(\tau)=\tau\left(1-\Theta\left(1-\mathrm{e}^{-b \tau}\right)\right)$ for $\Theta=0.6,0.8$, $0.9,0.95,1$ and $b=0.1$.

### 1.7.3.1 Case of "small" $\Theta$

We take now $\Theta \in\left[0, \Theta^{*}\right)$ and examine first whether the inverse of $V_{\Theta, b}$ exists.
Lemma 1.7.2. For $\Theta \in\left[0, \Theta^{*}\right)$, function $V_{\Theta, b}$ is strictly increasing, thus, $V_{\Theta, b}^{-1}$ exists and is strictly increasing on $(0,+\infty)$.

Proof. To show that function $V_{\Theta, b}$ is monotone, we calculate its derivative with respect to $\tau$ :

$$
V_{\Theta, b}^{\prime}(\tau)=\frac{\mathrm{d}}{\mathrm{~d} \tau}\left(\tau\left(1-\Theta\left(1-\mathrm{e}^{-b \tau}\right)\right)\right)=(1-\Theta)+\Theta \mathrm{e}^{-b \tau}(1-b \tau) .
$$

We now determine its zeros:

$$
\begin{aligned}
V_{\Theta, b}^{\prime}(\tau) & =0 \\
(1-\Theta)+\Theta \mathrm{e}^{-b \tau}(1-b \tau) & =0 \\
\mathrm{e}^{-b \tau}(1-b \tau) & =\frac{\Theta-1}{\Theta} \\
\mathrm{e}^{1-b \tau}(1-b \tau) & =\mathrm{e} \frac{\Theta-1}{\Theta} .
\end{aligned}
$$

With the notations $x:=1-b \tau$ and $\rho:=\mathrm{e}(\Theta-1) / \Theta<0$, we need to examine the solutions $x$ to the equation (1.44). The relation $\Theta<\Theta^{*}$ implies $\rho<-1$ /e, hence, there is no solution $x$ to equation (1.44) according to Remark 1.7.1. Thus, there are no zeros of function $V_{\Theta, b}^{\prime}$, therefore $V_{\Theta, b}$ is monotone. Furthermore, $V_{\Theta, b}^{\prime}(1 / b)=1-\Theta>0$ implies that function $V_{\Theta, b}$ is increasing on $(0,+\infty)$. Hence, its inverse $V_{\Theta, b}^{-1}$ exists and is strictly increasing on $(0,+\infty)$.

We state now the result for the non-negativity preservation.

Proposition 1.7.7. For $\Theta \in\left[0, \Theta^{*}\right)$, the non-negativity property $D_{1}$ holds for the method (1.59) if the time step $\tau$ satisfies the following criterion:

$$
\begin{equation*}
\tau \leq V_{\Theta, b}^{-1}\left(\frac{1}{M}\right) \tag{1.60}
\end{equation*}
$$

For the proof, see 33 .

### 1.7.3.2 Case of "large" $\Theta$

We take now $\Theta \in\left[\Theta^{*}, 1\right]$, and examine the behaviour of function $V_{\Theta, b}$.
Lemma 1.7.3. With Notations 1.7 .1 and 1.7 .3 , we have the following assertions.

1. For $\Theta \in\left[\Theta^{*}, 1\right)$, we have the following strictly monotonicity segments of function $V_{\Theta, b}$ :
(a) on $\left(0, \tau_{0}\right)$ the function $V_{\Theta, b}$ is strictly increasing, therefore, its inverse $V_{1}^{-1}$ exists and is strictly increasing,
(b) on $\left(\tau_{0}, \tau_{-1}\right)$ the function $V_{\Theta, b}$ is strictly decreasing, therefore, its inverse $V_{2}^{-1}$ exists and is strictly decreasing,
(c) on $\left(\tau_{-1},+\infty\right)$ the function $V_{\Theta, b}$ is strictly increasing, therefore, its inverse $V_{3}^{-1}$ exists and is strictly increasing
2. For $\Theta=1$, the inverse of function $V_{1, b}=\tau \mathrm{e}^{-b \tau}$ is strictly increasing on $[0,1 / b)$ and decreasing on $(1 / b,+\infty)$.

For the proof, see [104].
We have then the following result for the non-negativity property in this case.

Proposition 1.7.8. For $\Theta \in\left[\Theta^{*}, 1\right]$, the non-negativity property $D_{1}$ is fulfilled for method (1.59) in the following cases:

1. for $\frac{1}{M} \in\left(0, V_{\Theta, b}\left(\tau_{-1}\right)\right]$ : if $\tau \leq V_{1}^{-1}\left(\frac{1}{M}\right)$,
2. for $\frac{1}{M} \in\left(V_{\Theta, b}\left(\tau_{-1}\right), V_{\Theta, b}\left(\tau_{0}\right)\right]$ : if $\tau \in\left(0, V_{1}^{-1}\left(\frac{1}{M}\right)\right]$ or $\tau \in\left[V_{2}^{-1}\left(\frac{1}{M}\right), V_{3}^{-1}\left(\frac{1}{M}\right)\right)$,
3. for $\frac{1}{M}>V_{\Theta, b}\left(\tau_{0}\right)$ : if $\tau<V_{3}^{-1}\left(\frac{1}{M}\right)$
with Notations 1.7.1 and 1.7.3.

Proof. The non-negativity of $I_{k, \ell}^{n+1}$ and $R_{k, \ell}^{n+1}$ follows from the non-negativity of $S_{k, \ell}^{n+1}$, so we only have to prove only the latter one. Similarly as in the proof of Proposition 1.7.7, we need to determine the intervals where

$$
V_{\Theta, b}(\tau):=\tau\left(1-\Theta\left(1-\mathrm{e}^{-b \tau}\right)\right) \leq \frac{1}{T_{k, \ell}^{n}}
$$

To do so, we need an inverse of function $V_{\Theta, b}$ which has the three branches presented in Lemma 1.7.3. Property $D_{2}$ implies the estimate $T_{k, \ell}^{n} \leq M$ which provides the assertions.

Remark 1.7.3. As we have already pointed out, the cases $\Theta=0$ and $\Theta=1$ correspond to the sequential splitting methods (1.57) and (1.53), respectively. Since $V_{1, b}(\tau)=\tau$, its inverse $V_{1, b}^{-1}$ is the identity in 1.60, cf. Proposition 1.7.5. Furthermore, $V_{0, b}(\tau)=\tau \mathrm{e}^{-b \tau}$ implies $V_{0, b}^{-1}(y)=-W(-b y) / b$ having the two branches $W=W_{-1}$ and $W=W_{0}$ as in Proposition 1.7.2. Hence, as expected, the corresponding results in Propositions 1.7.7 and 1.7 .8 meet the conditions in Propositions 1.7 .2 and 1.7 .5 .

Remark 1.7.4. As in the case of the sequential splitting 1-2 (cf. Remark 1.7.2), Propositions 1.7 .7 and 1.7 .8 yield sharper conditions of the time step than the bound is when analysing the positivity preservation of the sub-problems separately. This is true due to the following consideration. The weighted splitting consists of the two sequential splittings 1-2 (1.53) and 2-1 1.57). In Proposition 1.7 .3 we showed that the bound obtained for sequential splitting $1-2$ is sharper than $1 / M$. Furthermore, for the sequential splitting $2-1$ we have the bound $1 / M$. Hence, the separate treatment of the sub-problems leads to the constraint $\tau \leq 1 / M$. The bounds obtained in the present section, however, are always sharper. In the case of "small" $\Theta$, the bound $V_{\Theta, b}^{-1}\left(\frac{1}{M}\right)$ is larger than $1 / M$, since $V_{\Theta, b}^{-1}(\tau)$ is a strictly increasing function in $\tau$ by Lemma 1.7.2, and $V_{\Theta, b}^{-1}(\tau)=\tau$ holds for $\Theta=0$ (so the bound here is simply $1 / M$ ), furthermore, $V_{\Theta, b}^{-1}(\tau)$ is monotonically decreasing as $\Theta$ increases. For "large" values of $\Theta$, we refer to Figure 1.7 to see that we can have three cases with respect to the location of the graph of function $V_{\Theta, b}(\tau)$ relative to the horizontal line $1 / M$. In all cases we have $\partial V_{\Theta, b}(\tau) / \partial \Theta<0$, that is, the left point of intersection moves to the right when the value of $\Theta$ is increasing (the movement might not be continuous but monotone). This means that the left bound is increasing as well.

### 1.7.4 Strang splitting

In contrast to the sequential splittings presented in Section 1.7.2 and 1.7.3. Strang splitting needs three steps with the two sub-problems Sub.1-Sub.2): the first step uses (Sub.1) with time step $\tau / 2$, the second uses (Sub.2) with time step $\tau$, and the third uses Sub.1) again with time step $\tau / 2$, always using the previous solution as an initial condition. Moreover, while the sequential splitting is of first order, the Strang is a second-order method, therefore we need to use a second-order time discretization method to avoid order reduction [32]. Hence, sub-problem (Sub.2) will be solved by Heun's method as presented in (1.42)-(1.43).

We note that the choice of (Sub.2) being the middle step is explained by the fact that it needs more computational effort and time than sub-problem (Sub.1). Hence, computing it only once at each step is more efficient than using the approach having (Sub.1) in the middle, since in that case (Sub.2) should be evaluated twice.

The corresponding steps to be solved one after another have then the following form with given $S_{k, \ell}^{n}, I_{k, \ell}^{n}, R_{k, \ell}^{n}$ values:

$$
\begin{gather*}
\left\{\begin{array}{l}
S_{k, \ell}^{[1], n+1}=\mathrm{e}^{-c \frac{\tau}{2}} S_{k, \ell}^{n}, \\
I_{k, \ell}^{[1], n+1}=\mathrm{e}^{-b \frac{\tau}{2}} I_{k, \ell}^{n}, \\
R_{k, \ell}^{[1], n+1}=R_{k, \ell}^{n}+\left(1-\mathrm{e}^{-c \frac{\tau}{2}}\right) S_{k, \ell}^{n}+\left(1-\mathrm{e}^{-b \frac{\tau}{2}}\right) I_{k, \ell}^{n},
\end{array}\right.  \tag{1.61}\\
\left\{\begin{array}{l}
\widehat{S}_{k, \ell}^{[2], n+1}=S_{k, \ell}^{[1], n+1}\left(1-\tau T_{k, \ell}^{[1], n+1}\right), \\
\widehat{I}_{k, \ell}^{[2], n+1}=I_{k, \ell}^{[1], n+1}+\tau S_{k, \ell}^{[1], n+1} T_{k, \ell}^{[1], n+1}, \\
\widehat{R}_{k, \ell}^{[2], n+1}=R_{k, \ell}^{[1], n+1},
\end{array}\right.  \tag{1.62}\\
\left\{\begin{array}{l}
S_{k, \ell}^{[2], n+1}=\frac{1}{2} S_{k, \ell}^{[1], n+1}+\frac{1}{2} \widehat{S}_{k, \ell}^{[2], n+1}\left(1-\tau \widehat{T}_{k, \ell}^{[2], n+1}\right), \\
I_{k, \ell}^{[2], n+1}=\frac{1}{2} I_{k, \ell}^{[1], n+1}+\frac{1}{2}\left(\widehat{I}_{k, \ell}^{[2], n+1}+\tau \widehat{S}_{k, \ell}^{[2], n+1} \widehat{T}_{k, \ell}^{[2], n+1}\right), \\
R_{k, \ell}^{[2], n+1}=R_{k, \ell}^{[1], n+1},
\end{array}\right.  \tag{1.63}\\
\left\{\begin{array}{l}
S_{k, \ell}^{n+1}=\mathrm{e}^{-c \frac{\tau}{2}} S_{k, \ell}^{[2], n+1}, \\
I_{k, \ell}^{n+1}=\mathrm{e}^{-b \frac{\tau}{2}} I_{k, \ell}^{[2], n+1}, \\
R_{k, \ell}^{n+1}=R_{k, \ell}^{[2], n+1}+\left(1-\mathrm{e}^{-c \frac{\tau}{2}}\right) S_{k, \ell}^{[2], n+1}+\left(1-\mathrm{e}^{-b \frac{\tau}{2}}\right) I_{k, \ell}^{[2], n+1}
\end{array}\right. \tag{1.64}
\end{gather*}
$$

for all $n \in \mathbb{N}$ and $\left(x_{k}, y_{\ell}\right) \in \mathcal{G}$. Since the form of the combined method would be too complex, we leave the steps individually written, and will study them separately. In this
case we get constraints which are more transparent and easier to verify. As before, we are going to show the validity of the properties $D_{2}-D_{4}$.

Proposition 1.7.9. Proposition 1.7.1 holds for the method (1.61)-(1.64).
Proof. It suffices to show the assertions for each steps (1.61)-1.64) by taking into account that they are constitutive steps, that is, their solution serves as an initial value for the next step.

The conservation of the size of the total population can be shown by adding up the equations in each step (1.61-(1.64). Since it is conserved in each step, it remains the same for the whole method as well.

By assuming the non-negativity property $D_{1}$ and using $\mathrm{e}^{-b \frac{\tau}{2}}, \mathrm{e}^{-c \frac{\tau}{2}} \in(0,1)$, we have that $S_{k, \ell}^{[1], n+1} \leq S_{k, \ell}^{n}$ and $R_{k, \ell}^{[1], n+1} \geq R_{k, \ell}^{n}$ in (1.61). Moreover, we have $T_{k, \ell}^{[1], n+1} \geq 0$ which, together with property $D_{1}$ for (1.62), implies $\widehat{S}_{k, \ell}^{[2], n+1} \leq S_{k, \ell}^{[1], n+1}$ and $\widehat{R}_{k, \ell}^{[2], n+1} \geq R_{k, \ell}^{[1], n+1}$ in (1.62). Again, the non-negativity of $I_{k, \ell}^{[2], n+1}$ implies $T_{k, \ell}^{[2], n+1}$, therefore, property $D_{1}$ holds for step (1.63), too. Since $\mathrm{e}^{-b \frac{\tau}{2}}, \mathrm{e}^{-c \frac{\tau}{2}} \in(0,1)$, properties $D_{3}$ and $D_{4}$ follow for step (1.64) as well.

Hence, as before, it suffices to analyse the conditions under which the non-negativity holds. The only difference from the previous sections is that in this case we will perform the analysis separately for each step (1.61)-(1.64). We will take into account, however, that they are constitutive steps of the method.

Proposition 1.7.10. The non-negativity property $D_{1}$ holds for the method (1.61)-(1.64) if

1. $\tau \in \mathbb{T}_{c / 2, M}$ for $2 M<b \mathrm{e}$,
2. $\tau \in \mathbb{T}_{p, M}$ with $p=\min \{b, c\} / 2$ for $2 M \geq$ be.

For the proof, see [33].
We remark that if the effect of the vaccination is not taken into account $(c=0)$, we have the condition $\tau<1 / M$, according to Lemma 1.7 .1 with $q=M$. This means that in this case we cannot guarantee a better sufficient condition on the time step than the one without applying operator splitting procedure, cf. (1.46).

### 1.7.5 Numerical results

The present section is devoted to the numerical illustration of our previously obtained theoretical results regarding (i) the preservation of the total density, (ii) the non-negativity of $S, I, R$, and (iii) the monotonicity of $S, R$.

There are issues already mentioned earlier which become really important at this point. Since the rectangular domain $\Omega$ is bounded, a special attention should be given to the boundary. As pointed out in Section 1.4, we assume that there is no susceptible population outside $\Omega$, thus, we assign zero values there. Using either a uniform or a non-uniform cubature, the cubature points usually do not belong to the spatial grid $\mathcal{G}$. To implement the cubature points at the boundary and in the corners as well, we define ghost cells outside the domain $\Omega$ having zero values. This enables the correct calculation of the values which correspond to the cubature points lying outside the domain.

For the numerical experiments, we use the same functions $g_{1}$ and $g_{2}$ which were defined in (1.6.3), namely

$$
\begin{aligned}
& g_{1}(r)=a(-r+\delta), \\
& g_{2}(\vartheta)=\beta \sin (\vartheta+\alpha)+\beta,
\end{aligned}
$$

where $a>0$ is the infection rate. We use the parameter values $\alpha=0$ and $\beta=1$ describing a northern wind on the domain. In our numerical experiments we take $a=100, b=0.1$, and $\delta=0.05$. We also use the Gaussian quadrature 1.22 defined in 1.4 .

Regarding the initial conditions, we assume that there are no recovered individuals at the beginning, that is, $R_{k, \ell}^{0}=0$ for all $\left(x_{k}, y_{\ell}\right) \in \mathcal{G}$. For the infected individuals, we use a Gaussian distribution concentrated at the middle of the domain $\left(\mathcal{L}_{1} / 2, \mathcal{L}_{2} / 2\right)$ which has a standard deviation $\tilde{\sigma}=\min \left(\mathcal{L}_{1}, \mathcal{L}_{2}\right) / 10$ :

$$
I_{k, \ell}^{0}=\frac{1}{2 \pi \tilde{\sigma}^{2}} \exp \left(-\frac{1}{2}\left[\left(\frac{h_{x}(k-1)-\frac{A}{2}}{\tilde{\sigma}}\right)^{2}+\left(\frac{h_{y}(\ell-1)-\frac{B}{2}}{\tilde{\sigma}}\right)^{2}\right]\right)
$$

where $\mathcal{L}_{1}=\left(P_{1}-1\right) h_{x}$ and $\mathcal{L}_{2}=\left(P_{2}-1\right) h_{y}$ as introduced in Section 1.4. We set here $\mathcal{L}_{1}=\mathcal{L}_{2}=1$. Due to property $D_{2}$, the sum of all individuals is constant in time at each point $\left(x_{k}, y_{\ell}\right) \in \mathcal{G}$. Thus, the initial distribution of the susceptibles can be chosen to be $S_{k, \ell}^{0}=N_{k, \ell}-I_{k, \ell}^{0}$ with some constant $N_{k, \ell}$ describing the sum of the species at point
$\left(x_{k}, y_{\ell}\right)$. For our tests, we choose $N_{k, \ell}=20$ for all $\left(x_{k}, y_{\ell}\right) \in \mathcal{G}$.
In Figure 1.8 the numerical solution is plotted for different time levels ( $S_{k, \ell}$ is plotted in the left column, $I_{k, \ell}$ in the middle, and $R_{k, \ell}$ on the right). One can see that the number of susceptibles decrease, and the number of infected moves towards the boundaries forming a wave. Both of them tend to the zero function, while the number $R_{k, \ell}$ of recovered tends to $N_{k, \ell}=20$ at each grid points $\left(x_{k}, y_{\ell}\right) \in \mathcal{G}$.

### 1.7.5.1 Testing the time step bounds

A natural question might arise on the strictness of the time step bounds derived in Section 1.7.2. Since Proposition 1.7 .1 holds for all schemes presented, it suffices to analyse the constraints on the non-negativity preservation. We note that the aforementioned choice of the parameters yields $M \approx 2.0893$.
a) Sequential splitting 1-2 (1.53)

Due to Proposition 1.7.2, the sufficient upper bound (the lower end of the forbidden interval) is

$$
\begin{equation*}
\widehat{\tau}=-\frac{1}{b} W_{0}\left(-\frac{b}{M}\right)=-\frac{1}{0.1} W_{0}\left(-\frac{0.1}{2.0893}\right) \approx 0.5033, \tag{1.65}
\end{equation*}
$$

while the sufficient lower bound (the upper end of the forbidden interval) is

$$
\widetilde{\tau}=-\frac{1}{b} W_{1}\left(-\frac{b}{M}\right)=-\frac{1}{0.1} W_{1}\left(-\frac{0.1}{2.0893}\right) \approx 45.5583
$$

In Table 1.4 we present our results on the time steps where the non-negativity property $D_{1}$ is preserved by the sequential splitting (1.53). In the second row we indicate the ratios $\tau / \widehat{\tau}$ (for small $\tau$ ) and $\tau / \widetilde{\tau}$ (for large $\tau$ ). The deficiency means the maximum of the absolute values of the negative values appeared in the solution at the final time level.

One can see that the necessary bound $\widehat{\tau}$ is relatively close to the numerically obtained "exact" bound. Moreover, certain errors appear when the time step is further increased, i.e., the solution becomes negative. It is also evident that after increasing the time step close enough to the other bound $\widetilde{\tau}$, the non-negativity property is satisfied again.
b) Sequential splitting $2-\mathbf{1} 1.57$


Figure 1.8: The numerical solutions $S_{k, \ell}^{n}, I_{k, \ell}^{n}, R_{k, \ell}^{n}$ shown in columns, respectively, at time levels $t=0, t=5, t=10, t=15, t=30$, for the sequential splitting (1.53).

In Proposition 1.7.5 we have the bound

$$
\begin{equation*}
\widehat{\tau}_{21}=\frac{1}{M} \approx \frac{1}{2.0893} \approx 0.4763 \tag{1.66}
\end{equation*}
$$

Table 1.4: Numerical results for sequential splitting 1-2 (1.53) for various time steps $\tau$. The deficiency is computed at final time $t_{f}=50$ for the upper, and at final time $t_{f}=400$ for the lower table.

| Time step $\tau$ | 0.48 | 0.50 | 0.52 | 0.54 | 0.56 | 0.58 | 0.60 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Ratio $\tau / \widehat{\tau}$ | 0.95 | 0.99 | 1.03 | 1.07 | 1.11 | 1.15 | 1.19 |
| Property $D_{1}$ | yes | yes | yes | yes | yes | no | no |
| Deficiency | 0 | 0 | 0 | 0 | 0 | 4.17e-3 | $2.42 \mathrm{e}-2$ |
| Time step $\tau$ |  |  | 37 | 40 | 43 | 46 |  |
| Ratio $\tau / \widetilde{\tau}$ |  |  | 0.81 | 0.88 | 0.94 | 1.01 |  |
| Property $D_{1}$ |  |  | no | no | yes | yes |  |
| Deficiency |  |  | $9.98 \mathrm{e}-1$ | 8.91e-2 | 20 | 0 |  |

Table 1.5 shows whether the non-negativity $D_{1}$ is preserved. The numerical experiments show that the behaviour of this method is similar to the previous one, although it produces slightly bigger errors. Also, it does not become stable for any bigger values of $\tau$, as expected from Proposition 1.7.5

Table 1.5: Numerical results for sequential splitting 2-1 1.57 for various time steps $\tau$. The deficiency is computed at final time $t_{f}=50$.

| Time step $\tau$ | 0.47 | 0.49 | 0.51 | 0.53 | 0.55 | 0.57 | 0.59 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Ratio $\tau / \widehat{\tau}_{21}$ | 0.98 | 1.03 | 1.07 | 1.11 | 1.15 | 1.19 | 1.24 |
| Property $D_{1}$ | yes | yes | yes | yes | no | no | no |
| Deficiency | 0 | 0 | 0 | 0 | $1.75 \mathrm{e}-3$ | $5.01 \mathrm{e}-3$ | $3.99 \mathrm{e}-2$ |

c) Weighted sequential splitting (1.59)

We study first the behaviour of the method for $\Theta=0.5<\Theta^{*}$. Proposition 1.7.7 leads to the bound

$$
\begin{equation*}
\widehat{\tau}_{w 1}=V_{\Theta, b}^{-1}\left(\frac{1}{M}\right) \approx V_{0.5,0.1}^{-1}\left(\frac{1}{2.0893}\right) \approx 0.4809 \tag{1.67}
\end{equation*}
$$

which is between the two previously obtained values (1.65) and 1.66). The corresponding errors are also between the errors of the two previous methods, which can be seen in Table 1.6

Next we study the case $\Theta=0.9>\Theta^{*}$. Then we get the following bounds from

Table 1.6: Numerical results for the weighted sequential splitting 1.59 with $\Theta=0.5$ for various time steps $\tau$. The deficiency is computed at final time $t_{f}=50$.

| Time step $\tau$ | 0.47 | 0.49 | 0.51 | 0.53 | 0.55 | 0.57 | 0.59 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Ratio $\tau / \widehat{\tau}_{w 1}$ | 0.98 | 1.02 | 1.06 | 1.10 | 1.14 | 1.18 | 1.23 |
| Property $D_{1}$ | yes | yes | yes | yes | no | no | no |
| Deficiency | 0 | 0 | 0 | 0 | $9.37 \mathrm{e}-4$ | $5.00 \mathrm{e}-3$ | $3.70 \mathrm{e}-2$ |

Proposition 1.7.8

$$
\begin{gathered}
\tau_{-1}=\frac{1}{b}\left(1-W_{-1}\left(\frac{\mathrm{e}(\Theta-1)}{\Theta}\right)\right)=\frac{1}{0.1}\left(1-W_{-1}\left(\frac{\mathrm{e}(0.9-1)}{0.9}\right)\right) \approx 27.6587, \\
\tau_{0}=\frac{1}{b}\left(1-W_{0}\left(\frac{\mathrm{e}(\Theta-1)}{\Theta}\right)\right)=\frac{1}{0.1}\left(1-W_{0}\left(\frac{\mathrm{e}(0.9-1)}{0.9}\right)\right) \approx 14.9596 .
\end{gathered}
$$

Since we have $1 / M \approx 0.4763$, being smaller than both of the above values, we have case (1) in Proposition 1.7.8. Therefore, we need to compute the following bound:

$$
\widehat{\tau}_{w 2}=V_{1}^{-1}\left(\frac{1}{M}\right) \approx V_{1}^{-1}\left(\frac{1}{2.0893}\right) \approx 0.5006,
$$

which is closer to the bound 1.65 than to (1.66). The corresponding results are listed in Table 1.7

Table 1.7: Numerical results for the sequential weighted splitting 1.59 with $\Theta=0.9$ for various time steps $\tau$. The deficiency is computed at final time $t_{f}=50$.

| Time step $\tau$ | 0.47 | 0.49 | 0.51 | 0.53 | 0.55 | 0.57 | 0.59 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Ratio $\tau / \widehat{\tau}_{w 2}$ | 0.93 | 0.98 | 1.02 | 1.06 | 1.10 | 1.14 | 1.18 |
| Property $D_{1}$ | yes | yes | yes | yes | yes | no | no |
| Deficiency | 0 | 0 | 0 | 0 | 0 | $1.46 \mathrm{e}-3$ | $5.63 \mathrm{e}-3$ |

d) Strang splitting (1.61)-1.64

By the choice of parameters, we have

$$
2 M=4.1786>0.2718=b e .
$$

Hence, we consider case (2) of Proposition 1.7.10. Moreover, relation $c=0.01<0.1=b$
leads to the bounds

$$
\begin{align*}
& \widehat{\tau}_{S}=-\frac{2}{c} W_{0}\left(-\frac{c}{2 M}\right) \approx-\frac{2}{0.01} W_{0}\left(-\frac{0.01}{4.1786}\right) \approx 0.4798,  \tag{1.68}\\
& \widetilde{\tau}_{S}=-\frac{2}{c} W_{1}\left(-\frac{c}{2 M}\right) \approx-\frac{2}{0.01} W_{1}\left(-\frac{0.01}{4.1786}\right) \approx 1626 . \tag{1.69}
\end{align*}
$$

As we can see, bound (1.68) is similar to the previously observed bounds 1.65, (1.66), and 1.67). Due to our choice of parameters, any recognizable dynamics of $S, I, R$ is already over before time level $t=1626$. Therefore, $\widetilde{\tau}_{S}$ in (1.69) is far too large to be considered as a suitable time step. Hence, we omit the numerical experiments using it. The numerical results are shown in Table 1.8.

Table 1.8: Numerical results for the Strang splitting (1.61)- 1.64 for various time steps $\tau$. The deficiency is computed at final time $t=50$.

| Time step $\tau$ | 0.47 | 0.49 | 0.51 | 0.53 | 0.55 | 0.57 | 0.59 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Ratio $\tau / \widehat{\tau}_{S}$ | 0.98 | 1.02 | 1.06 | 1.10 | 1.15 | 1.19 | 1.23 |
| Property $D_{1}$ | yes | yes | yes | yes | no | no | no |
| Deficiency | 0 | 0 | 0 | 0 | $5.39 \mathrm{e}-4$ | $7.30 \mathrm{e}-3$ | $2.71 \mathrm{e}-2$ |

### 1.7.5.2 Accuracy analysis

Besides the preservation of the qualitative properties, we also studied the accuracy of the presented methods. Since the exact solution to system (1.3) in not known, we considered a reference solution instead which was computed with a small time step. The time step was first chosen to be the bound acquired in the previous sections, and then by halving it six times, we got seven solutions. The last one was chosen to be the reference solution. We had a spatial mesh of $20 \times 20$ points and a bilinear interpolation with a $5 \times 5$ Gaussian quadrature, and the parameters $a=100, b=0.1, c=0.01$, and $\delta=0.1$. We chose the final time $t_{f}=20$. We define the relative global error at time level $t_{f}=N \tau$ as

$$
\varepsilon(\tau):=\frac{\left\|\underline{X}^{N}-X^{N}\right\|}{\left\|\underline{X}^{N}\right\|}
$$

with $X \in\{S, I, R\}$, where $X^{N}$ means the matrix with elements $X_{k, \ell}^{N}$ for $\left(x_{k}, y_{\ell}\right) \in \mathcal{G}$, and the underlying refers to the reference solution. We took the relative global error by using the maximum norm and the discrete $L^{1}$ and $L^{2}$ norms.

In Figure 1.9 the order plot can be seen for the relative errors $\varepsilon(\tau)$ of the four presented
splitting schemes, where the colours represent the methods and the line styles the various norm types: the solid line states for the maximum norm, the dashed line for the discrete $L^{1}$ norm, and the dotted line for the discrete $L^{2}$ norm. Since the slope of the curve on the $\log$-log plot corresponds to the order of the method, one can see that the sequential and the weighted splittings are of first order, while the Strang splitting is of second order convergent. We fit a line to the data and obtained the values given in Table 1.9. The fairly good approximations to the theoretical orders can be clearly read from the data.


Figure 1.9: Order plot for the relative global error of the four splitting schemes identified with colours. Solid line states for the maximum norm, dashed line for the discrete $\mathrm{L}^{1}$ norm, and dotted line for the discrete $L^{2}$ norm

Table 1.9: Slopes of the lines fitted to the curves in Figure 1.9 which approximate the orders of the methods in the three norms considered.

| splitting | order in max. norm | order in $L^{1}$ norm | order in $L^{2}$ norm |
| :--- | :---: | :---: | :---: |
| Sequential 1-2 $(1.53)$ | 1.11046 | 1.11216 | 1.11263 |
| Sequential 2-1 (1.57) | 1.10598 | 1.10145 | 1.09238 |
| Weighted (1.59) with $\Theta=0.9$ | 1.04794 | 1.04965 | 1.05635 |
| Strang (1.61) 1.64 | 1.97442 | 1.98131 | 1.97148 |

### 1.8 Conclusions

In this chapter we observed an integro-differential equation, which can be used to model space-dependent epidemic processes. Our main aim was to construct such numerical schemes which do not only have high order, but also possess the discrete analogues of the properties of the original continuous equation.

First it was proved that the aforementioned equation (1.3) has a unique solution with (biologically) reasonable properties. Later the original partial integro-differential equation was transformed to an integral form and a numerical solution was also proposed to solve it, but it turned out that the method is only of first order accurate, so its application is not advised.

Then, another method was observed involving semi-discretization in space, and then several different time discretizations were proposed, including the first order (forward and backward) methods of Euler, higher order Runge-Kutta methods and also several different operator splitting techniques. It turns out that for a sufficiently chosen time step, these methods preserve the properties of the original continuous model.

A possible extension of this model involves the introduction of an incubation period, which results in a delay integro-differential equation. This model is described in the next chapter. Other possible extensions are mentioned in Chapter 3

## Chapter 2

## Epidemic models with latency period

The aim of this chapter is to extend the models which have already been studied in Chapter 1. Some diseases take some time to develop inside an infected individual, so they do not start to infect upon their infection but only after a short period of time (so-called latency period). For this, we generalize the previous models by using delay integro-differential equations. In this chapter we show that the solution of the continuous problems possesses these features and give sufficient conditions that guarantee the properties to the numerical solutions.

Delay can be introduced into the space-dependent SIR model described in Chapter 1 in several ways, which arise from different considerations. Here we will only examine the ones resulting in a constant delay, which were also studied thoroughly by several authors [28, 75, 114]. The key difference here from the previous models of Chapter 1 is the fact that the infected group $I$ does not only contain those who are infectious, but also those who are infected, but cannot yet transmit the disease. This phenomenon might happen when the disease needs some time to develop in a host, so during this time the infected individual is not infectious. Consequently, the size of group $I$ might change not upon meetings of individuals coming from the groups of $S$ and $I$, but rather ones from $S$ and an infectious one from $I$. However, since the illness takes a constant time to develop (let us denote this constant value from now on as $\sigma$ ), the number of infectious individuals at time $t$ at place $(x, y)$ is $I(t-\sigma, x, y)$. Also, the processes of the recovery or the vaccination is not changed, so that part of the equation remains the same as in (1.3).

Note that instead of collecting all of the infected people in group $I$, one can introduce another group $E$ usually called exposed, which contains all of the individuals who are infected, but cannot yet infect others. These models are usually referred to as SEIR models, but since this modification would considerably increase the number of equations which have to be solved numerically, we will not discuss them here.

By the previous arguments, the equation describing our new process can be written
in the form

$$
\left\{\begin{array}{l}
\frac{\partial S(t, x, y)}{\partial t}=-S(t, x, y) \mathcal{F}(t-\sigma, x, y)-c S(t, x, y)  \tag{2.1}\\
\frac{\partial I(t, x, y)}{\partial t}=S(t, x, y) \mathcal{F}(t-\sigma, x, y)-b I(t, x, y) \\
\frac{\partial R(t, x, y)}{\partial t}=b I(t, x, y)+c S(t, x, y)
\end{array}\right.
$$

in which $\mathcal{F}$ is defined as in 1.8).
Because (2.1) is a delay system, to obtain a properly posed problem we also need the values of the function $I$ on the time interval $[-\sigma, 0]$. This is the so-called history function that will be denoted by $I_{h}:[-\sigma, 0] \times \Omega \rightarrow \mathbb{R},(t, x, y) \mapsto I_{h}(t, x, y)$. Later we will show that some assumptions are needed on the history function to assure that the solution is continuous and behaves as expected. The history functions of $S$ ans $R$ are denoted similarly but these functions do not appear in the model.

The model does not have boundary condition in a classical sense but due to the integral in 1.8 we assume that $I$ is equal to zero outside the domain $\Omega$.

The structure of this chapter is as follows. In Section 2.1 we prove that the system (2.1) has a unique solution, and in Section 2.1.1 it is shown that this solution possesses some biologically reasonable properties. In Section 2.2.1 we discretize the system (2.1) in space, and then in Sections 2.2 .2 and 2.2 .3 different time discretizations are examined, namely the method of Euler and Runge-Kutta methods, respectively. The corresponding numerical experiments can be found in Section 2.3.

### 2.1 Existence and properties of the analytic solution

In this section we discuss the solvability of system (2.1). The key tool will be the method of steps introduced by Bellman [16], which involves the splitting of our time interval [ $0, t_{f}$ ] into smaller intervals with length $\sigma$ with the grid points $0, \sigma, 2 \sigma, 3 \sigma, \ldots, t_{f}$. The ( $n+1$ )th element of this list will be denoted by $t_{n}$. Let us denote the solution on the time interval $\left(t_{n-1}, t_{n}\right]$ by $S_{n}(t, x, y), I_{n}(t, x, y)$ and $R_{n}(t, x, y)$. Then the delay differential equation (2.1) on a given interval $\left(t_{n}, t_{n+1}\right](n \geq 1)$ becomes a classical partial integro-differential
equation

$$
\left\{\begin{array}{l}
\frac{\partial S_{n+1}(t, x, y)}{\partial t}=-S_{n+1}(t, x, y) \mathcal{F}_{n}(t-\sigma, x, y)-c S_{n+1}(t, x, y)  \tag{2.2}\\
\frac{\partial I_{n+1}(t, x, y)}{\partial t}=S_{n+1}(t, x, y) \mathcal{F}_{n}(t-\sigma, x, y)-b I_{n+1}(t, x, y) \\
\frac{\partial R_{n+1}(t, x, y)}{\partial t}=b I_{n+1}(t, x, y)+c S_{n+1}(t, x, y)
\end{array}\right.
$$

with initial conditions

$$
S_{n+1}\left(t_{n}, x, y\right)=S_{n}\left(t_{n}, x, y\right), \quad I_{n+1}\left(t_{n}, x, y\right)=I_{n}\left(t_{n}, x, y\right), \quad R_{n+1}\left(t_{n}, x, y\right)=R_{n}\left(t_{n}, x, y\right)
$$

and the notation

$$
\begin{equation*}
\mathcal{F}_{n}(t-\sigma, x, y)=\mathcal{F}\left(I_{n}(t-\sigma, x, y)\right)=\int_{0}^{\delta} \int_{0}^{2 \pi} g_{1}(r) g_{2}(\theta) I_{n}(t-\sigma, \bar{x}(r, \theta), \bar{y}(r, \theta)) r \mathrm{~d} \theta \mathrm{~d} r . \tag{2.3}
\end{equation*}
$$

In the case of $t \in(0, \sigma],(2.2)$ has the form

$$
\left\{\begin{array}{l}
\frac{\partial S_{1}(t, x, y)}{\partial t}=-S_{1}(t, x, y) F_{h}(t-\sigma, x, y)-c S_{1}(t, x, y)  \tag{2.4}\\
\frac{\partial I_{1}(t, x, y)}{\partial t}=S_{1}(t, x, y) F_{h}(t-\sigma, x, y)-b I_{1}(t, x, y) \\
\frac{\partial R_{1}(t, x, y)}{\partial t}=b I_{1}(t, x, y)+c S_{1}(t, x, y)
\end{array}\right.
$$

with initial conditions

$$
S_{1}(0, x, y)=S_{h}(0, x, y), \quad I_{1}(0, x, y)=I_{h}(0, x, y), \quad R_{1}(0, x, y)=R_{h}(0, x, y)
$$

and the notation

$$
\begin{equation*}
\mathcal{F}_{h}(t-\sigma, x, y)=\mathcal{F}\left(I_{h}(t-\sigma, x, y)\right)=\int_{0}^{\delta} \int_{0}^{2 \pi} g_{1}(r) g_{2}(\theta) I_{h}(t-\sigma, \bar{x}(r, \theta), \bar{y}(r, \theta)) r \mathrm{~d} \theta \mathrm{~d} r . \tag{2.5}
\end{equation*}
$$

Because the structures of (2.2) and (2.4) are the same, we will only consider the case of (2.2) but all the statements also will be valid for (2.4).

If we proceed to solve the equation step by step (i.e. solving it on $(0, \sigma]$ then on $(\sigma, 2 \sigma]$ and so on) then the term $F_{n}(t-\sigma, x, y)$ is known, since it only depends on $I_{n}(t, x, y)$. It is also important to notice that the first equation of $(2.2)$ does not depend on the
later ones, thus can be solved independently of the others. After this $S_{n+1}(t, x, y)$ will be known, which means that the only unknown is $I_{n+1}(t, x, y)$ in the second equation, so it can also be solved directly. Finally, the last equation can also be solved by integration, resulting in the solution of system (2.2).

For the sake of simplicity, we introduce the function $\tilde{\mathcal{F}}_{n+1}(t, x, y):=\mathcal{F}_{n}(t-\sigma, x, y)$. In other words, we shift the domain of $\mathcal{F}_{n+1}(t-\sigma, x, y)$ from $\left(t_{n-1}, t_{n}\right] \times \Omega$ to $\left(t_{n}, t_{n+1}\right] \times \Omega$. Then (2.2) has the form $\left(t \in\left(t_{n}, t_{n+1}\right]\right)$

$$
\left\{\begin{array}{l}
\frac{\partial S_{n+1}(t, x, y)}{\partial t}=-S_{n+1}(t, x, y) \tilde{\mathcal{F}}_{n+1}(t, x, y)-c S_{n+1}(t, x, y)  \tag{2.6}\\
\frac{\partial I_{n+1}(t, x, y)}{\partial t}=S_{n+1}(t, x, y) \tilde{\mathcal{F}}_{n+1}(t, x, y)-b I_{n+1}(t, x, y) \\
\frac{\partial R_{n+1}(t, x, y)}{\partial t}=b I_{n+1}(t, x, y)+c S_{n+1}(t, x, y)
\end{array}\right.
$$

Theorem 2.1.1. Assume that the history functions $S_{h}(t, x, y), I_{h}(t, x, y)$ and $R_{h}(t, x, y)$ are continuous in time. Then system (2.1) has a unique solution, which is continuously differentiable in time on $\left(0, t_{f}\right]$.

For the proof, see [105].
In the next section we observe whether the continuous solution possesses some biological features.

### 2.1.1 Qualitative properties of the analytic solution

A mathematical model is considered to be reasonable not only when the system has only one solution but the behavior of such solution should also possess the properties of the biological model. Here we are going to observe four of these properties, namely the ones which were already defined in Section 1.2. Our goal is to show that the properties $C_{1}-$ $C_{4}$ are satisfied by the solution of system (2.1).

Theorem 2.1.2. Suppose that properties $C_{1}-C_{4}$ hold for our history functions, which are also continuous. Then $C_{1}-C_{4}$ also hold for the solution of system (2.1) on a time interval $\left(0, t_{f}\right]$.

For the proof, see [105].
In the next sections we examine the semi-discretized and the fully discretized versions of (2.1) and check whether their solutions satisfy the analogous versions of $C_{1}-C_{4}$.

### 2.2 Numerical approximation

### 2.2.1 The spatially discretized models and their properties

It is evident that Bellman's method can only be used in practice when the integral of the history function $I_{h}$ at the time $t=-\sigma$ is known, which is usually not the case. Because of this, in the following sections we propose a numerical approach to approximate these analytic solutions.

Upon looking at system (2.1), it is evident that the most problematic part is the fact that it contains integrals on its right-hand side. We are taking a similar path as in Section 1.4. first we approximate the integral terms inside our equation, and then discretize it in space using a grid defined in our domain.

We define some two-dimensional cubature formula on the disc $B_{\delta}(x, y)$ with positive weights to approximate the integral $\mathcal{F}((t, x, y))$. As in Section 1.4 . let us denote by $\mathcal{Q}(x, y)$ the set of cubature nodes in the disk $B_{\delta}(x, y)$ parametrized by polar coordinates, i.e.,

$$
\mathcal{Q}(x, y):=\left\{\left(x_{i j}, y_{i j}\right)=\left(x+r_{i} \cos \left(\theta_{j}\right), y+r_{i} \sin \left(\theta_{j}\right)\right) \in B_{\delta}(x, y), i \in \mathcal{I}, j \in \mathcal{J}\right\},
$$

where $r_{i}$ denotes the distance from center point $(x, y), \theta_{j}$ is the angle, and $\mathcal{I}$ and $\mathcal{J}$ are the set of indices of cubature nodes. Using numerical integration, we get the system

$$
\left\{\begin{array}{l}
\frac{\partial S(t, x, y)}{\partial t}=-S(t, x, y) T(t-\sigma, \mathcal{Q}(x, y))-c S(t, x, y)  \tag{2.7}\\
\frac{\partial I(t, x, y)}{\partial t}=S(t, x, y) T(t-\sigma, \mathcal{Q}(x, y))-b I(t, x, y) \\
\frac{\partial R(t, x, y)}{\partial t}=b I(t, x, y)+c S(t, x, y)
\end{array}\right.
$$

where

$$
T(t-\sigma, \mathcal{Q}(x, y))=\sum_{\left(x_{i j}, y_{i j}\right) \in \mathcal{Q}(x, y)} w_{i, j} g_{1}\left(r_{i}\right) g_{2}\left(\theta_{j}\right) I\left(t-\sigma, x+r_{i} \cos \left(\theta_{j}\right), y+r_{i} \sin \left(\theta_{j}\right)\right),
$$

(when $t-\sigma \in[-\sigma, 0)$, we take $I_{h}$ in the definition), and $w_{i, j}>0$ are the weights of the cubature formula.

It is clear that the arguments detailed in Section 2.1 can be used similarly, which
results in the following theorem.

Theorem 2.2.1. Assume that the history functions are continuous and properties $C_{1}-C_{4}$ hold for them. Then system (2.7) has a unique, continuously differentiable solution, which also has properties $C_{1}-C_{4}$.

A natural question is the choice of the numerical approximation of the integral. In Section 1.4 two separate choices of cubatures were investigated. One of them, the ElhayKautsky cubature results in a uniform distribution of points on the unit disc, while the other, the Gauss-Legendre cubature (which involves a transformation of the integral to a unit square) results in a non-uniform distribution. Numerical experiments show that while the first one works well for polynomials, the second one is better for arbitrary nonlinear functions. Since we cannot guarantee that our function $I(t, x, y)$ is a polynomial, we are going to use here the latter one. For further details of the different methods, see [104.

### 2.2.1.1 The properties of the semi-discretized model

As in Section 1.4 , we assume that our domain $\Omega$ is a rectangle with one vertex at the origin, i.e. $\Omega=\left(0, \mathcal{L}_{1}\right) \times\left(0, \mathcal{L}_{2}\right), \mathcal{L}_{1}, \mathcal{L}_{2} \in \mathbb{R}^{+}$. Note that the following arguments also hold for domains in a more general form, but involve much more careful choice of spatial grids.

Let us discretize our rectangle shaped domain using the spatial grid

$$
\mathcal{G}=\left\{\left(x_{k}, y_{\ell}\right) \in \Omega \mid x_{k}=(k-1) h_{1}, y_{\ell}=(\ell-1) h_{2}, 1 \leq k \leq P_{1}, 1 \leq \ell \leq P_{2}\right\},
$$

supposing that $\left(P_{1}-1\right) h_{1}=\mathcal{L}_{1}$ and $\left(P_{2}-1\right) h_{2}=\mathcal{L}_{1}$. This grid consists of $P_{1} P_{2}$ points with spatial step sizes $h_{1}$ and $h_{2}$, and we approximate the continuous solutions by a matrix containing the values at these grid points.

After this semi-discretization, we get the following set of equations:

$$
\left\{\begin{array}{l}
\frac{d S_{k, \ell}(t)}{d t}=-S_{k, \ell}(t) T_{k, \ell}\left(t-\sigma, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right)-c S_{k, \ell}(t)  \tag{2.8}\\
\frac{d I_{k, \ell}(t)}{d t}=S_{k, \ell}(t) T_{k, \ell}\left(t-\sigma, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right)-b I_{k, \ell}(t) \\
\frac{d R_{k, \ell}(t)}{d t}=b I_{k, \ell}(t)+c S_{k, \ell}(t)
\end{array}\right.
$$

in which $X_{k, \ell}(t)(X \in\{S, I, R\})$ denotes the approximation of the function at grid point
$\left(x_{k}, y_{\ell}\right)$ and at time $t$ and

$$
\begin{equation*}
T_{k, \ell}\left(t-\sigma, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right):=\sum_{\left(\bar{x}_{k}, \bar{y}_{l}\right) \in \mathcal{Q}\left(x_{k}, y_{\ell}\right)} w_{i, j} g_{1}\left(r_{i}\right) g_{2}\left(\theta_{j}\right) \hat{I}_{k, \ell}\left(t-\sigma, \bar{x}_{k}, \bar{y}_{\ell}\right), \tag{2.9}
\end{equation*}
$$

where $\bar{x}_{k}=x_{k}+r_{i} \cos \left(\theta_{j}\right)$ and $\bar{y}_{l}=y_{\ell}+r_{i} \sin \left(\theta_{j}\right)\left(\right.$ when $t-\sigma<0$, we have $\left(I_{h}\right)_{k, \ell}$ in the formula).

Note that the points ( $\bar{x}_{k}, \bar{y}_{k}$ ) might not be part of $\mathcal{G}$, so there are no $I_{k, \ell}$ values assigned to them. Because of this, we will approximate them by some interpolation method using the nearest known $I_{k, \ell}$ values and positive coefficients. (This is the reason for the $\hat{I}$ notation.) In order to satisfy the qualitative properties, it is important to choose such interpolations that preserve non-negativity in the sense that if the known values are nonnegative, then the function we get at the end of our process should also be non-negative. Such interpolations include monotone interpolation that uses piecewise cubic Hermite interpolating polynomials [41, 49] ('pchip' for short), which will be used in the numerical experiments.

As in the previous section, the methods described in Section 2.1 can be used again for system (2.8), which results in the following theorem.

Theorem 2.2.2. Assume that the history functions are continuous and properties $C_{1}-C_{4}$ hold for them. Then system (2.8) has a unique, continuously differentiable solution, which also has properties $C_{1}-C_{4}$.

In Sections 2.2.2 and 2.2.3 we present two different numerical methods for system (2.8): first we solve it using the explicit Euler method via the Elsgolts approach [43], and later positivity-preserving Runge-Kutta methods [92, 93, 94 .

### 2.2.2 Time discretization: explicit Euler method

One of the key elements in the solution of delay differential equations is the fact that the discontinuities should be included in the mesh of the time discretization. Since our history function is smooth, the only discontinuities in the higher order derivatives might appear at the points $k \sigma, k \in \mathbb{N}$. Based on this, we define them as

$$
\mathcal{G}_{t}=\left\{t_{n / q} \in\left[-\sigma, t_{f}\right] \left\lvert\, t_{n / q}=n \frac{\sigma}{m}\right., n \in \mathbb{Z},-q \leq n \leq q \frac{t_{f}}{\sigma},\right\}
$$

where $q$ is a positive integer.

On this above mesh, we can define the scheme

$$
\left\{\begin{array}{l}
S^{n+1}=S^{n}-\tau S^{n} \circ T^{n-q}-c \tau S^{n},  \tag{2.10}\\
I^{n+1}=I^{n}+\tau S^{n} \circ T^{n-q}-b \tau I^{n}, \\
R^{n+1}=R^{n}+b \tau I^{n}+c \tau S^{n},
\end{array}\right.
$$

where $0 \leq n \leq q \frac{t_{f}}{\sigma}$ and $\tau=\frac{\sigma}{q}$. The symbol $\circ$ is the element-by-element or Hadamard product of the matrices and the $(k, \ell)$ element of the matrix $X^{n}(X \in\{S, I, R\})$ is the approximation of the value $X_{k, \ell}\left(t_{n / q}\right)$, moreover the $(k, \ell)$ element of $T^{n-q}$ gives the approximation of $T\left(t_{n / q}-\sigma, \mathcal{Q}\left(x_{k}, y_{\ell}\right)\right)$ by interpolating the elements of $I^{n-q}$.

Instead of analyzing the previous numerical method in terms of its convergence, we are going to observe how well the model describes the real-life processes: more precisely, whether our model preserves the discrete versions of qualitative properties $C_{1}-C_{4}$. Similarly as before, we denote these by $D_{1}-D_{4}$.

Now we prove that for a sufficiently small time step (or in other words, a sufficiently large $m$ ) properties $D_{1}-D_{4}$ hold for $n>0$.

Theorem 2.2.3. Suppose that properties $D_{1}-D_{4}$ hold for the history functions discretized on the grids $\mathcal{G}$ and $\mathcal{G}_{t}$, then property $D_{2}$ holds without restrictions. Furthermore, if we assume that the time step satisfies

$$
\begin{equation*}
\tau=\frac{\sigma}{q} \leq \min \left\{\frac{1}{\widehat{T}+c}, \frac{1}{b}\right\} \tag{2.11}
\end{equation*}
$$

where $\widehat{T}$ is given by (1.27), then properties $D_{1}, D_{3}$ and $D_{4}$ also hold up to the step $n \leq q \frac{t_{f}}{\sigma}$.

For the proof, see [105].
An important remark is that the step size must be in the form $\frac{\sigma}{q}$, which means that by condition (2.11), the theoretically best step size is in the form $\frac{\sigma}{\widetilde{q}}$, where

$$
\widetilde{q}=\min \left\{q \in \mathbb{N}^{+} \left\lvert\, \frac{\sigma}{q}<\min \left\{\frac{1}{\widehat{T}+c}, \frac{1}{b}\right\}\right.\right\} .
$$

### 2.2.3 Time discretization: Application of Runge-Kutta methods

To achieve higher order convergence in our numerical approximation, we need to use higher order methods. One of the most widely used ones are the Runge-Kutta methods. Since in this chapter we consider a constant delay, these methods are easily applicable.

In the sequel, the investigations will be based on the canonical Shu-Osher forms which were introduced in Section 1.6.1. If these are applied to the initial value problem

$$
\left\{\begin{array}{l}
U^{\prime}(t)=\mathcal{P}(U(t))  \tag{2.12}\\
U(0)=U_{0}
\end{array}\right.
$$

in which $\mathcal{P}$ is the operator describing the discretization in space, and $U_{0}$ is a given function, then the corresponding Runge-Kutta method had the form

$$
\begin{align*}
U^{(i)} & =v_{i} U^{n-1}+\sum_{j=1}^{m} \alpha_{i j}\left(U^{(j)}+\frac{\tau}{p} \mathcal{P}\left(U^{(j)}\right)\right), \quad 1 \leq i \leq m+1  \tag{2.13}\\
U^{n} & =U^{(m+1)}
\end{align*}
$$

in which $U^{n}=\left(U_{1}^{n}, \ldots, U_{N}^{n}\right)$ is the numerical solution at time $t_{n}$ and $U^{(i)}=\left(U_{1}^{(i)}, \ldots, U_{N}^{(i)}\right)$ is the approximation of the solution at the $i$ th stage of the method. Also, $p$ is a parameter describing the relation between the coefficients of the Butcher form.

In our case, the problem $\sqrt{1.16}$ ) can be written in the form

$$
\begin{equation*}
U^{\prime}(t)=\mathcal{P}(U(t), U(t-\sigma)), \quad u \text { is given on }[-\sigma, 0] . \tag{2.14}
\end{equation*}
$$

Because of this, the explicit Runge-Kutta method applied to (2.14) takes the form

$$
\begin{align*}
U^{(i)} & =v_{i} U^{n-1}+\sum_{j=1}^{m} \alpha_{i j}\left(U^{(j)}+\frac{\tau}{\mathcal{C}} \mathcal{P}\left(U^{(j)}, U^{n-q}\right)\right), \quad 1 \leq i \leq m+1  \tag{2.15}\\
U^{n} & =U^{(m+1)}
\end{align*}
$$

( $0 \leq n \leq q t_{f} / \sigma$ ) and $\mathcal{C}$ is the corresponding SSP-coefficient (see Section 1.6.1).
The next theorem states that for a sufficiently small time step, properties $D_{1}-D_{4}$ hold for the above scheme.

Theorem 2.2.4. Consider an explicit Runge-Kutta method in the form (2.15) with SSP
coefficient $\mathcal{C}>0$ and applied to system (2.8) with non-negative history function. Then $D_{2}$ holds, and the properties $D_{1}, D_{3}$ and $D_{4}$ also hold if the step-size satisfies

$$
\begin{equation*}
\tau=\frac{\sigma}{m} \leq \mathcal{C} \min \left\{\frac{1}{\widehat{T}+c}, \frac{1}{b}\right\} \tag{2.16}
\end{equation*}
$$

where $\widehat{T}$ is given by (1.27).
For the proof, see [105].
We should also note here that like in the case of the explicit Euler method, we cannot use arbitrary values for time steps, but they should be in the form $\sigma / q$. Therefore, the theoretically best time step is $\sigma / \widetilde{q}$, where

$$
\widetilde{q}=\min \left\{q \in \mathbb{N}^{+} \left\lvert\, \frac{\sigma}{q}<\mathcal{C} \min \left\{\frac{1}{\widehat{T}+c}, \frac{1}{b}\right\}\right.\right\}
$$

### 2.3 Numerical experiments

In this section we present some numerical experiments to confirm our previous results. First we show that the bounds we got in Theorems 2.2 .3 and 2.2 .4 are sharp in the sense that the use of bigger time steps results in qualitatively bad behavior. Then, in the second part we present some graphs on which the solutions are compared for different values of $\sigma$ and their qualitative properties are checked.

### 2.3.1 Construction of the test problem

The equation (2.1) is solved on the rectangle domain $\Omega=\left(0, \mathcal{L}_{1}\right) \times\left(0, \mathcal{L}_{2}\right)\left(\mathcal{L}_{1}, \mathcal{L}_{2} \in \mathbb{R}^{+}\right)$. We set the parameters as $\mathcal{L}_{1}=\mathcal{L}_{2}=1$ and $c=0.01$. In order to be able to define the term $\mathcal{F}(t-\sigma, x, y)$, we choose the functions

$$
g_{1}(r)=a(\delta-r) \quad \text { and } \quad g_{2}(\theta)=1
$$

(Here a constant $g_{2}$ function is used, since in the case of epidemics it is a reasonable assumption.)

In the numerical experiments, the choice $a=100$ is used. The radius $\delta$, delayparameter $\sigma$ and the rate of recovery $b$ will be set later. The history functions are chosen

$$
\left\{\begin{array}{l}
S_{h}(t, x, y)=20-I_{h}(t, x, y)  \tag{2.17}\\
I_{h}(t, x, y)=\frac{1}{2 \pi \tilde{\sigma}^{2}} \exp \left(-\frac{1}{2}\left[\left(\frac{x-\frac{1}{2}}{\tilde{\sigma}}\right)^{2}+\left(\frac{y-\frac{1}{2}}{\tilde{\sigma}}\right)^{2}\right]\right)\left(1+\frac{t}{\sigma}\right), \\
R_{h}(t, x, y)=0
\end{array}\right.
$$

for $t \in[-\sigma, 0]$, where $I_{h}(t, x, y)$ is a scaled Gaussian distribution with standard deviation $\tilde{\sigma}=1 / 10$ concentrated at the center $(1 / 2,1 / 2)$ of the domain $\Omega$. Note that since $I_{h}$ is a monotone increasing continuous non-negative function and $1 /\left(2 \pi \tilde{\sigma}^{2}\right) \approx 15.92<20$, functions 2.17) fulfill properties $C_{1}-C_{4}$. The graphs of the history functions at time $t=0$ can be seen in Figure 2.1.


Figure 2.1: The history functions $S_{h}, I_{h}, R_{h}$ at $t=0$ shown in columns, respectively.

The semi-discretization is carried out on a standard rectangular mesh with step sizes $h_{1}=h_{2}=1 / 19$. As mentioned before, we can use different cubatures to approximate the integral $\mathcal{F}(I(t, x, y))$ (see [104]) - here we are going to use a $40 \times 40$ Gaussian one which was already described in Sections 1.4 and 1.6.3. Therefore, the cubature has the form

$$
\begin{aligned}
& T(t-\sigma, x, y)= \\
&=\sum_{i=1}^{40} \sum_{j=1}^{40} w_{i} w_{j} a\left(-\xi_{j} \delta+\delta\right) I\left(t-\sigma, x+\xi_{j} \delta \cos \left(2 \pi \eta_{l}\right), y+\xi_{j} \delta \sin \left(2 \pi \eta_{l}\right)\right) \xi_{i} 2 \pi \delta^{2}=, \\
&=\sum_{k=1}^{40^{2}} \widetilde{w}_{k} a\left(-r_{m}+\delta\right) I\left(t-\sigma, x+\widetilde{x}_{m}, y+\widetilde{y}_{m}\right),
\end{aligned}
$$

where $m=40(i-1)+j, \widetilde{x}_{m}=\xi_{i} \delta \cos \left(2 \pi \eta_{j}\right), \widetilde{y}_{m}=\xi_{i} \delta \sin \left(2 \pi \eta_{j}\right), \widetilde{w}_{m}=w_{i} w_{j} 2 \pi \delta^{2} \xi_{i}$ and $r_{m}=\sqrt{\left(x+x_{m}\right)^{2}+\left(y+y_{m}\right)^{2}}$. Based on this, at the given spatial grid point $\left(x_{k}, y_{\ell}\right)$, the
approximation

$$
T_{k, \ell}\left(t-\sigma, x_{k}, y_{\ell}\right)=\sum_{i=1}^{40^{2}} \widetilde{w}_{k} a\left(-\widetilde{r}_{m}+\delta\right) \hat{I}\left(t-\sigma, x_{k}+\widetilde{x}_{m}, y_{\ell}+\widetilde{y}_{m}\right)
$$

is used, where $\widetilde{r}_{m}=\sqrt{\left(x_{k}+x_{m}\right)^{2}+\left(y_{\ell}+y_{m}\right)^{2}}$ and $\hat{I}\left(t-\sigma, x_{k}+\widetilde{x}_{m}, y_{\ell}+\widetilde{y}_{m}\right)$ is computed using piecewise cubic Hermite interpolation.

With the help of the previous constructions, the time discretization methods discussed in Sections 2.2.2 and 2.2.3 now can be applied (see the next subsections).

### 2.3.2 Sharpness of the time step bounds

In the previous sections, namely in Theorems 2.2.3 and 2.2.4 we gave sufficient conditions for the qualitatively good behavior of the numerical solution, i.e. if we use a smaller time step than the bounds, then our numerical solution possesses properties $D_{1}-D_{4}$. A natural question which might arise in the case of such sufficient conditions is the effect of the use of bigger time steps.

In Table 2.1, we can see the theoretical bound (2.11)

$$
\begin{equation*}
\tau \leq \min \left\{\frac{1}{\left(20 \sum_{i=1}^{40} \sum_{j=1}^{40} w_{i} w_{j} 100 \delta^{3}\left(1-\xi_{i}\right) \xi_{i} 2 \pi\right)+0.01}, 10\right\} \tag{2.18}
\end{equation*}
$$

denoted by "theor. b.", the actual time step in the form $\sigma / \widetilde{q}$ (see the end of Section 2.2.2) denoted by "time step", and the bound calculated by experiments, i.e. the time step in the form $\sigma / q_{\text {exp }}$ for which the method works as expected but for $\sigma /\left(q_{\text {exp }}-1\right)$ it gives qualitatively inaccurate results - this value is denoted by "real b." in the table. Also, the difference $q_{\text {exp }}-\widetilde{q}$ is denoted by "diff." in the table. The bound can be considered sharp when this value is zero. The last column shows the ratio of the "time step" and the "real bound", i.e. the sharpness of the bound we got from our theorem. As it can be seen, this ratio is 1 for several parameter values.

We show an example for the qualitatively bad behaviour of the method and show the sharpness of the obtained time step bound (2.18) in the second row of Table 2.1. On the left panel of Figure 2.2, the numerical solution $S^{n}$ can be seen at the time level $t=3$ obtained with time step $\tau=\sigma / 4=1 / 4=0.25$. We can see that the solution is qualitatively correct, namely, the values are non-negative. But when we use the next

| $\delta$ | $\sigma$ | theor. b. | time step | real b. | diff. | ratio |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 0.13 | 1 | 0.2169 | 0.2 | 0.2 | 0 | 1.0000 |
| 0.12 | 1 | 0.2755 | 0.25 | 0.25 | 0 | 1.0000 |
| 0.15 | 0.3 | 0.1413 | 0.1 | 0.1 | 0 | 1.0000 |
| 0.15 | 0.5 | 0.1413 | 0.125 | 0.125 | 0 | 1.0000 |
| 0.14 | 0.4 | 0.1737 | 0.1333 | 0.1333 | 0 | 1.0000 |
| 0.13 | 0.5 | 0.2169 | 0.1667 | 0.1667 | 0 | 1.0000 |

Table 2.1: Numerical results for the explicit Euler method 2.10 for various time steps with final time $t_{f}=15$ and $b=0.05$ (the other parameters are given before).
possible time step $\tau=\sigma / 3=1 / 3=0.3333$, we also get negative values. On the right panel of Figure 2.2 the white area corresponds to those grid points at which the solution is negative. Thus, the obtained bound is sharp.


Figure 2.2: The numerical solution $S^{n}$ shown at final time $\mathcal{T}=3$ computed with time steps $\tau=0.25$ (left) and $\tau=0.333$ (right), with parameters $\delta=0.12$ and $\sigma=1$. The white area corresponds to those grid points at which the solution becomes negative.

The results for the second order Runge-Kutta method are presented in Table 2.2. Here we can see that the theoretical bound is not that sharp in all of the cases, so our condition is only sufficient, but not necessary. However, in some cases the theory gives time steps that are not far from the best possible one (when we have small numbers in the 'diff.' column). For different choices of the parameters and initial conditions, we might get even better results, resulting in a sharp bound even in the higher order case.

| $\delta$ | $\sigma$ | $b$ | theor. b. | time step | real b. | diff. | ratio |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 0.13 | 1 | 0.1 | 0.2169 | 0.2000 | 0.5000 | 3 | 0.4000 |
| 0.12 | 1 | 0.1 | 0.2755 | 0.2500 | 0.5000 | 2 | 0.5000 |
| 0.13 | 0.5 | 0.05 | 0.2169 | 0.1667 | 0.2500 | 1 | 0.6667 |
| 0.135 | 0.5 | 0.05 | 0.1937 | 0.1667 | 0.2500 | 1 | 0.6667 |
| 0.135 | 0.4 | 0.01 | 0.1937 | 0.1333 | 0.2000 | 1 | 0.6667 |

Table 2.2: Numerical results for the RK2 method 2.10 for various time steps with final time $\mathcal{T}=15$.

### 2.3.3 Comparison of the cases with different delay parameters

In this section, we present some graphs of the numerical solutions of equation (2.1) with different values of $\sigma$. We are plotting the solution at final time $t_{f}=7$ with parameters $b=0.1$ and $\delta=0.1$, computed with the possible largest time step below the theoretical bound 0.4752 (computed similarly as 2.18 with $\mathcal{C}=1$ ) and the second order RungeKutta method is used. As we can see in Figure 2.3, the increase of parameter $\sigma$ results in a slower spread of the infection, which corresponds to the biological requirements (a longer latent period results in a slower pandemic).

### 2.4 Conclusions

In this chapter we examined space-dependent epidemic models in which we assumed that there is a latency period, i.e. it takes some time for infected individuals to be infectious. This resulted in a system of delay integro-differential equations. The main aim of the chapter was the construction of such numerical methods which preserve the qualitative properties of the original continuous equation.

In Section 2.1, we proved that the aforementioned delay integro-differential equation has a unique solution. Moreover, this solution possesses biologically reasonable properties. Later, in Section 2.2.1 the integral terms of the equations were approximated with some quadrature, and then using a spatial discretization we got a system of delay differential equations.

The next sections described the time discretizations which can be used to solve the aforementioned system of delay differential equations numerically. Since we only consider the case of a constant delay, the usual numerical schemes used to solve ordinary differential equations can be applied: in Section 2.2 .2 the method of Euler, while in Section 2.15 the


Figure 2.3: The numerical solutions $S^{n}, I^{n}, R^{n}$ at $t_{f}=7$ shown in columns, respectively, $\sigma$ values $\sigma=0.2, \sigma=0.5, \sigma=1.0$ and $\sigma=2.0$.
method of Runge-Kutta was observed. It turned out that by using a necessarily small time step, the numerical solution possesses biologically reasonable properties. Then, in Section 2.3 some numerical experiments were conducted, including the observation whether the obtained bounds are not only sufficient, but also necessary or not.

## Chapter 3

## Possible further extensions

In Chapters 1 and 2, a mathematical model describing epidemics propagating in space, and a space-dependent one with latency period are presented, respectively. In this chapter we mention some possible further extensions of these models.

Another way to describe epidemics is to consider a graph which could model the connections inside a population (with the nodes being the individuals), and then define an ordinary differential equation at each vertex modeling the possible three stages of the given individual. When we consider very large graphs, the connections might become nearly impossible to track down - in these cases, it is sometimes useful to consider not a very large graph, but a limit of such graphs in some sense [74]. Then, the equations previously defined on finite graphs become a system of partial differential equations defined on the limit object (a graphone), or more precisely on the square $[0,1] \times[0,1]$. The properties of such equations were examined in [66].

It is also important to realize that in the case of equations (1.15) the births and natural deaths are neglected in the domain, i.e. in such cases the total number of species at a given point is assumed to be constant in time. A possible extension of this model is to introduce some terms which describe these neglected phenomena. In these cases the proofs of Section 1.1 are very similar, and bounds similar to the ones in the later sections can also be given. An important difference though is that property $C_{1}$, and consequently $D_{1}$ is different in this case, since the number of species at a given point is not constant - however, if the rate of birth and natural deaths are constants in time, an ordinary differential equation can be given for the total number of species at a given point, which can be solved, so the total number of species at a given point can be computed. A similar argument can be used to extend the delay model (2.1).

An important remark regarding the models which were described before is that they ignore any diffusion or other movements inside the observed region. Such effects of diffusion can be introduced into the model by adding a Laplacian term to each of the equations,
which might have even non-integer powers. However, if we use spatial discretizations which approximate the Laplacian terms using linear combinations of the values of the function with positive coefficients, then similar bounds can be given as in the previous Chapters.

A possible extension of the delay model involves non-constant latency periods. In these cases the method of Elsgolts cannot be used, since the point $t_{n}-\sigma\left(t_{n}\right)$ might not be part of the mesh defined on the time interval. Because of this, the values of the solutions at these points are calculated using some interpolation. However, if we use a positivity preserving interpolation (like piecewise cubic Hermite interpolation 'pchip' which was mentioned in e.g. Section 2.3), then a similar bound can be proven like in the previous cases. Note though that the size of memory in these schemes is considerably larger, since one has to store all the values which were computed previously.

Another different approach is to discretize the space dependent or the delay systems in space using the method of finite elements. In these cases the use of positivity preserving elements results in similar bounds as discussed before.

## Part II

## Modeling the collapse of Easter Island

In recent years numerous papers and books have discussed the events that led to the demographic collapse on Easter Island (Rapa Nui) in the 16th and 17th century. When in 1786 comte de Lapérouse, French explorer stepped on the island as the first European, he found only 2000 people with a much less developed civilization than the one which would have be required to build such big monuments. Another unusual observation was the lack of trees on the island. In the next centuries, several papers and books tried to describe the events that could lead to the demographic collapse. Most of them blamed the irresponsible inhabitants and the reckless consumption of goods on the island. Some of them even claimed that these events could happen globally, so our increasing growth will lead to the fall of humanity (see [11). Since the increasing popularity of the concept of sustainable development, these theories have gained even more recognition.

In the early 2000s Terry Hunt, historian from the University of Hawai'i arrived on the island to confirm these theories [60, 61]. However, he found no traces of the long decline of economy proposed in the original theory. The collected data showed a shorter and much drastic collapse, which led Hunt to the realization that some other factors could have effects on the events. Because of the numerous rat corpses and chewed seeds, he proposed a new model involving the Polynesian rats (Rattus exulans). These animals could have been originally brought to the island by the settlers themselves (some theories even suppose that these animals were transported to Rapa Nui for food - this concept was studied in [90], but we will neglect this effect). However, because the rats ate the seeds of the trees, the reproduction of trees was decreased so dramatically, that the population of plants could not cope with the constant harvest done by the settlers.

The theories of Hunt were formalized by Basener et al. in three papers. First they examined the classical theories [13], then added the rats to it [14]. In their third work, they constructed a spatial model from the previous one, adding diffusion terms to the equations. On the ground of their surprising results, we added a new term to their equations, and examined the effect of this action. This model is described in Chapter 4. Sections 4.1.1 and 4.1.3 collect the results of 99, Section 4.1.2 summarizes 98 while Section 4.2 compiles the theories of [101]. In Chapter 5 we construct a two-dimensional model similar to the one in the previous chapter: the main statements of [102] are listed in this part. The last chapter of this part mentions some possible extensions of the presented models.

## Chapter 4

## One dimensional model

As mentioned before, we would like to model the ecological collapse that took place on Easter Island. For this, we use ordinary differential equations describing the changes of population on the island. The theories of Hunt were formalized by Basener et al. [14] in the following way:

$$
\begin{align*}
\frac{d P}{d t} & =a P\left(1-\frac{P}{T}\right) \\
\frac{d R}{d t} & =c R\left(1-\frac{R}{T}\right)  \tag{4.1}\\
\frac{d T}{d t} & =\frac{b}{1+f R} T\left(1-\frac{T}{M}\right)-h P
\end{align*}
$$

in which $P, R, T$ denote the number of people, rats and trees, and $a, b$ and $c$ are the birth ratio for humans, for trees and for rats, respectively. Also, $f$ is the destructive effect of the rats on trees, $M$ is the maximum amount of trees which can live on the island, and $h$ is the number of trees cut down by a person in a year.

Note that in these models $P=1$ means one person, $T=1$ means the number of trees used by a person in a year, and $R=1$ is the number of rats which could be supported by one tree (where one tree means the number of trees described above).

### 4.1 Discretized model

In [15], Basener et al. used the following spatial invasive species model to represent the dynamics on Easter Island: they thought of Rapa Nui as an island which has a volcano in the middle, so they split the habitable coast into $N$ regions and labeled them from 1 to $N$ (Figure 4.1). Therefore, the neighbours of region $s$ (if $1<s<N$ ) are the regions with labels $s-1$ and $s+1$. Also, the region with label 1 and the one with label $N$ are next to each other.


Figure 4.1: The split of the habitable coast into N regions.

They also introduced the $D_{p}$ and $D_{r}$ parameters to measure the constant diffusion of people and rats, respectively. In this way, they got the following equations:

$$
\begin{aligned}
\frac{d P^{s}}{d t} & =a P^{s}\left(1-\frac{P^{s}}{T^{s}}\right)+D_{p}\left(P^{s-1}-2 P^{s}+P^{s+1}\right) \\
\frac{d R^{s}}{d t} & =c R^{s}\left(1-\frac{R^{s}}{T^{s}}\right)+D_{r}\left(R^{s-1}-2 R^{s}+R^{s+1}\right) \\
\frac{d T^{s}}{d t} & =\frac{b}{1+f N R^{s}} T^{s}\left(1-\frac{T^{s}}{\frac{M}{N}}\right)-h P^{s}
\end{aligned}
$$

where $P^{s}, R^{s}$ and $T^{s}$ denote the number of people, rats and trees in region $s$ respectively $(s \in\{1, \ldots N\})$.

In their article, it turned out that upon increasing either the constant $D_{p}$ or $D_{r}$, the system becomes unstable. This was a surprising result since it is usually supposed in scientific models that diffusion has a stabilizing effect on the system. However, one can suppose that the source of this instability comes from the system's asymmetry: the first two equations involve diffusion, while the third one does not. In a closed system like Easter Island, the movement of seeds (by the wind or on the fur of the animals) cannot be overlooked. Also, in ecology it is a well known phenomenon that tropical trees tend to have descents far from themselves, which also shows the necessity of the addition of this term (this effect is described by the Janzen-Connell hypothesis [27, 64]).

In this way we can extend the previous model in the way presented below, in which
$D_{t}$ is the diffusion coefficient of trees mentioned before.

$$
\begin{align*}
\frac{d P^{s}}{d t} & =a P^{s}\left(1-\frac{P^{s}}{T^{s}}\right)+D_{p}\left(P^{s-1}-2 P^{s}+P^{s+1}\right) \\
\frac{d R^{s}}{d t} & =c R^{s}\left(1-\frac{R^{s}}{T^{s}}\right)+D_{r}\left(R^{s-1}-2 R^{s}+R^{s+1}\right)  \tag{4.2}\\
\frac{d T^{s}}{d t} & =\frac{b}{1+f N R^{s}} T^{s}\left(1-\frac{T^{s}}{\frac{M}{N}}\right)-h P^{s}+D_{t}\left(T^{s-1}-2 T^{s}+T^{s+1}\right)
\end{align*}
$$

An important remark here is that when $T^{s}=0$, the first two equations are not welldefined. A possible solution to this problem is presented in Section 4.1.4.

Apart from the trivial equilibriums (where one of the species die out), 4.2) will have a coexistence equilibrium, which is the same as computed in [14] and has the following form:

$$
\begin{equation*}
P_{\varepsilon}=R_{\varepsilon}=T_{\varepsilon}=\frac{1}{N} \frac{M(b-h)}{b+h M f} \tag{4.3}
\end{equation*}
$$

for every one of the $N$ systems.
If we linearize the system at equilibrium (4.3), we get the following form:

$$
\begin{array}{r}
\left(\begin{array}{c}
\frac{d P^{s}}{d t} \\
\frac{d R^{s}}{d t} \\
\frac{d T^{s}}{d t}
\end{array}\right)=\left[\begin{array}{ccc}
-a & 0 & a \\
0 & -c & c \\
-h & \frac{-f M h(b-h)}{b(1+f M)} & \frac{f M h-b+2 h}{1+f M}
\end{array}\right]\left(\begin{array}{c}
P^{s} \\
R^{s} \\
T^{s}
\end{array}\right)+  \tag{4.4}\\
+\left(\begin{array}{c}
D_{p}\left(P^{s-1}-2 P^{s}+P^{s+1}\right) \\
D_{r}\left(R^{s-1}-2 R^{s}+R^{s+1}\right) \\
D_{t}\left(T^{s-1}-2 T^{s}+T^{s+1}\right)
\end{array}\right)
\end{array}
$$

We decouple the equations using the Fourier transformation meaning that we transform the previous system of $3 N$-many equations with unknown functions $P^{s}(t), R^{s}(t)$ and $T^{s}(t)$ (where $s$ is the index of the given region) into a new system of 3 N -many equations with unknown functions $x_{q}(t), y_{q}(t)$ and $z_{q}(t)$ (where $q$ is the index of the new variables). Our main goal is to get rid of the connections between equations describing different
regions.
For the first two equations (with functions $x_{q}$ and $y_{q}$ ) we get the same results as [15], while the equation of the $z_{q}$ function can be computed in the following way:

$$
\begin{gather*}
\frac{d z_{q}}{d t}=\frac{1}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i q s}{N}} \frac{d T^{s}}{d t}= \\
=\frac{1}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i q s}{N}}  \tag{4.5}\\
{\left[-h P^{s}-\frac{f M h(b-h)}{b(1+f M)} R^{s}+\frac{f M h-b+2 h}{1+f M} T^{s}+D_{t}\left(T^{s-1}-2 T^{s}+T^{s+1}\right)\right] .}
\end{gather*}
$$

From now on, we will use the following notations

$$
\begin{aligned}
A & :=\frac{f M h(b-h)}{b(1+f M)}, \\
B & :=\frac{f M h-b+2 h}{1+f M} .
\end{aligned}
$$

With these, we can rewrite (4.5) as follows:

$$
\begin{gathered}
\frac{1}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i q s}{N}}\left[-h P^{s}-A R^{s}+B T^{s}+D_{t}\left(T^{s-1}-2 T^{s}+T^{s+1}\right)\right]= \\
=-h \frac{1}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i q s}{N}} P^{s}-A \frac{1}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i q s}{N}} R^{s}+B \frac{1}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i q s}{N}} T^{s}+ \\
+D_{t} \frac{1}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i q s}{N}} T^{s-1}-D_{t} \frac{2}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i q s}{N}} T^{s}+D_{t} \frac{1}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i q s}{N}} T^{s+1}= \\
=-h x_{q}-A y_{q}+B z_{q}+D_{t} e^{\frac{-2 \pi i q}{N}} \frac{1}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i r(s-i)}{N}} T^{s-1}-2 D_{t} z_{q}+ \\
+D_{t} e^{\frac{2 \pi i q}{N}} \frac{1}{N} \sum_{s=1}^{N} e^{\frac{-2 \pi i q(s+i)}{N}} T^{s+1}= \\
=-h x_{q}-A y_{q}+B z_{q}+D_{t} e^{\frac{-2 \pi i q}{N}} z_{q}-2 D_{t} z_{q}+D_{t} e^{\frac{2 \pi i q}{N}} z_{q}= \\
=-h x_{q}-A y_{q}+\left[B-2 D_{t}\left(1-\cos \frac{2 \pi q}{N}\right)\right] z_{q}= \\
=-h x_{q}-A y_{q}+\left[B-4 D_{t} \sin ^{2} \frac{\pi q}{N}\right] z_{q} .
\end{gathered}
$$

Thus, the decoupled system can be written as:

$$
\left(\begin{array}{c}
\frac{d x_{q}}{d t}  \tag{4.6}\\
\frac{d y_{q}}{d t} \\
\frac{d z_{q}}{d t}
\end{array}\right)=\left(\begin{array}{ccc}
-\left[a+4 D_{p} \sin ^{2} \frac{\pi q}{N}\right] & 0 & a \\
0 & -\left[c+4 D_{r} \sin ^{2} \frac{\pi q}{N}\right] & c \\
-h & -A & B-4 D_{t} \sin ^{2} \frac{\pi q}{N}
\end{array}\right)\left(\begin{array}{c}
x_{q} \\
y_{q} \\
z_{q}
\end{array}\right)
$$

If we use the values of the constants from [14], which were $a=0.03, b=1, c=10$, $M=12000$ and $h=0.25$, the matrix from equation (4.6) gets the following shape:

$$
S_{q}:=\left(\begin{array}{ccc}
-\left[0.03+4 D_{p} \sin ^{2} \frac{\pi q}{N}\right] & 0 & 0.03  \tag{4.7}\\
0 & -\left[10+4 D_{r} \sin ^{2} \frac{\pi q}{N}\right] & 10 \\
-0.25 & \frac{-2250 f}{1+12000 f} & \frac{6000 f-1}{2+24000 f}-4 D_{t} \sin ^{2} \frac{\pi q}{N}
\end{array}\right)
$$

For example, if we choose the values $f=0.001$ and $N=10$ (like in [15]), we get

$$
\left(\begin{array}{ccc}
-\left[0.03+4 D_{p} \sin ^{2} \frac{\pi q}{10}\right] & 0 & 0.03 \\
0 & -\left[10+4 D_{r} \sin ^{2} \frac{\pi q}{10}\right] & 10 \\
-0.25 & -\frac{9}{52} & \frac{5}{26}-4 D_{t} \sin ^{2} \frac{\pi q}{10}
\end{array}\right)
$$

for every $q=1, \ldots 10$ value.
The system is stable, if every matrix is stable. Thus, if we want to examine the stability of the system, we have to examine all the $N$ (in the previous example, ten) matrices. One region is stable, if all the eigenvalues of the corresponding matrix have a negative real part. If this property holds for every matrix, then the system is stable, otherwise it is unstable.

In the next section, we fix the $f$ and $N$ parameters and investigate the effect of the added tree diffusion to our system.

### 4.1.1 The effect of the diffusion of trees

As it was mentioned in the previous section, Basener et al. found in [15] that the increase of $D_{p}$ or $D_{r}$ makes the system (4.6) unstable. In this section we show that the increase
of $D_{t}$ has an opposite effect, such that it makes our system stable.
Instead of calculating the eigenvalues of our matrices, we use the Routh-Hurwitz criterion. For $3 \times 3$ matrices, the theorem can be formulated as follows.

Theorem 4.1.1 ([63, [89]). A matrix $S \in \mathbb{R}^{3 \times 3}$ is stable if and only if

- $\operatorname{det}(S)<0$,
- $\operatorname{tr}(S)<0$,
- $\operatorname{tr}(S) M_{2}(S)<\operatorname{det}(S)$,
where $\operatorname{det}(S)$ is the determinant of $S, \operatorname{tr}(S)$ is the trace of $S$ and $M_{2}(S)$ is the sum of the three $2 \times 2$ principal minors of $S$.

Theorem 4.1.2. Let us suppose that the model parameters satisfy the condition

$$
\begin{equation*}
B^{2}-A c-a h<0 . \tag{4.8}
\end{equation*}
$$

Then, if (4.7) is stable for $D_{t}=0$, then it is stable for all positive $D_{t}$ values.
For the proofs of this and the following theorems, see [99.
Remark 4.1.1. If the parameters satisfy the condition

$$
\max \left\{B^{2},\left(B-4 D_{t}^{*}\right)^{2}\right\}-A c-a h<0
$$

for a given $D_{t}^{*}>0$ value, then if the matrix in (4.6) is stable for $D_{t}^{*}$, then it is stable for every other $D_{t}>D_{t}^{*}$ values.

Remark 4.1.2. The condition (4.8) can be guaranteed for the values in (4.7) if we choose $f$ to be greater than

$$
-\frac{107}{2184000}+\frac{1}{291200} \sqrt{282} \approx 8.6751 \times 10^{-6}
$$

Theorem 4.1.3. Let us suppose that the parameters in (4.6) satisfy the conditions

$$
\begin{align*}
B & <\min \{a+c, A+h\},  \tag{4.9}\\
B\left(A c+a h+(a+c)^{2}\right) & <c^{2}(A+a)+B^{2}(a+c)+a^{2}(c+h) .
\end{align*}
$$

Then for fixed positive $D_{p}$ and $D_{r}$ values there is a positive $\overline{D_{t}}$ value for which the system (4.6) is stable for every $D_{t}>\overline{D_{t}}$ value.

Remark 4.1.3. The condition (4.9) is satisfied for the values in matrix (4.7), which means that for any fixed ( $D_{p}, D_{r}$ ) pair the matrix (4.7) is stable for sufficiently large $D_{t}$ values.

On the following pages, our goal will be to draw a three-dimensional surface which will be the border of the stable and the unstable $\left(D_{p}, D_{r}, D_{t}\right)$ triplets for the values $f=0.001$ and $N=10$ (we call a triplet stable if the corresponding system is stable, and call the triplet unstable if the corresponding system is unstable).

Instead of determine the points of the previously mentioned border precisely, we calculate them numerically using a bisection method in the following way. Let us assume that we examine one parameter (let us call it $x$ ) and we search for those values of $x$ at which our system changes its stability. In this case, we choose one point $\left(x_{0}\right)$ to be very small (in the algorithm, we use 0 ) and the other $\left(x_{1}\right)$ to be very large (we choose this to be $10^{7}$ ). If the system is stable in $x_{0}$ and unstable in $x_{1}$, then the border is in between these two, so we check the stability at point $\frac{x_{0}+x_{1}}{2}$. If it is stable, then the border is above this point, and if it is unstable, then it is under this point. Then we continue this iteration until the distance of $x_{0}$ and $x_{1}$ is small (we usually use 0.0001 ), and then we say that the border is the mean of the two endpoints of the last interval. Using this method, we get the graph for $D_{t}=0$ in Figure 4.2. On this graph, the horizontal axis is $D_{p}$ and the vertical axis is $D_{r}$. The stable points are below the graph, and the points on it and above are unstable. Note that we got the same results as [15]: when $D_{p}=0$, the system gets unstable for $D_{r}>0.15$ ( 0.155 in [15) and in the case of $D_{r}=0$ the same happens for $D_{p}>0.09$ (exactly the same as in [15]).

If we increase the $D_{t}$ value, we get the graph in Figure 4.3. As we can see, the area of stability gets larger as the diffusion of the trees increases. It can also be seen from further analysis that the system becomes stable if $D_{t}>0.41$ for every value of $D_{r}$ and $D_{p}$.

It can also give us important information if we examine those $D_{t}$ points for which the system changes stability at certain $\left(D_{p}, D_{r}\right)$ pairs. For this, we will plot the critical $D_{t}$ points as the function of $D_{p}$, and plot these borders for several $D_{r}$ values. When $D_{r} \rightarrow \infty$, these borders tend to a certain function, which can be seen in Figure 4.4

From these graphs we can suspect that for any $\left(D_{p}, D_{r}\right)$ pair our system is stable if $D_{t}$ is bigger than a certain value (probably around 0.51 ). These results also confirm the


Figure 4.2: The border of stability in the case $D_{t}=0$. The stable points are below the graph.


Figure 4.3: The increasing border of stability in the cases $D_{t}=\{0,0.015,0.03, \ldots 0.15\}$. The stable points are below the graph.
statement of Theorem 4.1.3.
We can also use a three-dimensional surface to represent the region of stability. The graph in Figure 4.5 is the border of the stability region. The vertical axis is $D_{r}$, the one going left is $D_{p}$ and the right one is $D_{t}$. We can see that as we increase $D_{t}$, the stability region gets bigger and bigger, until the system becomes stable for every $D_{r}$ and $D_{p}$ value. As we increase $D_{r}$, the system becomes unstable after a certain point, unless $D_{t}$ is bigger than the value 0.51 discussed above, because in this case the system does not change stability.

It is clear from the previous claims that there are certain $\left(D_{r}, D_{t}\right)$ pairs for which the change of the diffusion of the people will have no effect on the stability, because it is stable for every $D_{p}$ value. When $D_{t}$ and $D_{r}$ are small, then the increase of $D_{p}$ will


Figure 4.4: Upper: The critical $D_{t}$ points as the function of $D_{p}$. The graphs are plotted from bottom-up with the values $D_{r}=50 k^{2}, k \in\{1,2, \ldots 10\}$. Lower: the same functions zoomed in to the $[0,3]$ interval.


Figure 4.5: The border of stable $\left(D_{p}, D_{r}, D_{t}\right)$ triplets. The stable ones are below our surface, and the unstable ones are above it.
cause the system to become unstable. For example, if we look at the stability region for $D_{r}=0$, we get the graph in Figure 4.6, where the points above the graph are stable and under it are unstable.


Figure 4.6: The border of stable $\left(D_{p}, D_{t}\right)$ points for $D_{r}=0$. The stable points are above the graph.

We can also see that for some other $\left(D_{r}, D_{t}\right)$ pairs our system might remain stable (e.g for $D_{t}>0.51$ ). However, for certain $\left(D_{r}, D_{t}\right)$ values the diffusion of people has a stabilizing effect. For example if we look at the line of $\left(D_{r}, D_{t}\right)=(100,0.2)$, our system is unstable for $D_{p}=0$, will be stable for around $D_{p}=0.15$ and will be unstable again if $D_{p}>0.2$.

In conclusion, we can say that these numerical results support the theorems stated before. Also, the increase of parameter $D_{r}$ destabilizes our system, the increase of $D_{t}$ has the opposite effect, and $D_{p}$ can act both ways depending on the values of $D_{p}$ and $D_{r}$.

### 4.1.2 The choice of the parameters $f$ and $N$

As it was mentioned before, the values $f=0.001$ and $N=10$ were chosen in [15], and only the parameters of diffusion were changed (we also used the same method in the previous section). However, it is not clear how this choice affects the stability of the system. On the following pages, we will carry out this analysis: we fix the ( $D_{p}, D_{r}, D_{t}$ ) triplet and only change the parameters $f$ and $N$. In this way we will search for those $(f, N)$ pairs (for a fixed $\left(D_{p}, D_{r}, D_{t}\right)$ triplet) where our system is stable and those for which it is unstable.

From now on, we call an $(f, N)$ pair stable, if our system is stable at those values (with a fixed diffusion parameter triplet $\left(D_{p}, D_{r}, D_{t}\right)$ ). Similarly, we call it unstable if
our system is unstable for those $f$ and $N$ values. Therefore, we define the function $g: \mathbb{N} \backslash\{0\} \rightarrow \mathbb{R}$ in the following way:

$$
g(N):=f_{N} \in \mathbb{R} \quad \text { s.t. } \quad \forall \varepsilon>0,\left(f_{N}-\varepsilon, N\right) \text { is stable and }\left(f_{N}+\varepsilon, N\right) \text { is unstable. }
$$

In other words, we will search for that value of $f$ for a fixed $N$ where upon increasing $f$ the system becomes unstable. We will assume that there is at most one such value for every $N$ : numerical experiments show that this is the case. Note that it may happen that $g(N)<0$ or $g(N)=\infty$, which means that for some $\left(D_{p}, D_{r}, D_{t}\right)$ triplets the system will be either always unstable or always stable. From now on, we will examine those cases when $0 \leq g(N)<\infty$, because these are the interesting ones (these are the ones in which stability changes).

Proposition 4.1.1. For every fixed $\left(D_{p}, D_{r}, D_{t}\right)$ triplet

$$
g(N) \longrightarrow \inf _{n \in \mathbb{N}^{+}} g(n), \text { as } N \rightarrow \infty
$$

The convergence of $g(N)$ means that there is such a constant $N_{0}$ that if $N>N_{0}$, then the values of $g(N)$ are very close, meaning that the value $f$ where the system changes stability will be similar for every value of $N>N_{0}$. Also, since the form of the limit is known, if we know the value of $\inf _{n \in \mathbb{N}^{+}} g(n)$ and there is such a value $N_{\text {min }}$ for which $g\left(N_{\text {min }}\right)=\inf _{n \in \mathbb{N}^{+}} g(n)$, then in our examination we can choose the value of $N$ to be $N=N_{\text {min }}$. In our numerical experiments we got that in all cases $N_{\text {min }}=2$ (see Conjecture 4.1.4, meaning that the use a large $N$ is unnecessary since for large values of $N$ (which would mean the calculation of the eigenvalues of $N$-many matrices) we get the same behaviour as in the case of $N=2$. (Of course, a larger value of $N$ would mean a better approximation of the biological model so it is usually necessary.)

For the proof of Proposition 4.1.1, we define the following sequence for every value of $f$ :

$$
a_{f}(N):=\left\{\begin{array}{r}
1, \text { if the system is stable for }(f, N) \\
-1, \text { if the system is unstable for }(f, N)
\end{array}\right.
$$

Proposition 4.1.2. For every fixed value of $f \neq \inf _{n \in \mathbb{N}^{+}} g(n), a_{f}(N)$ converges as $N \rightarrow \infty$.

Note that since $a_{f}$ has finite range, the convergence means that

$$
\exists N_{\text {conv }}: \forall n, m>N_{\text {conv }}: a_{f}(n)=a_{f}(m) .
$$

For the proof of proposition 4.1.2, see [98]. Here we mention only the facts that in the proof it turns out that if $a_{f}(n)=-1$ for any value of $n$, then $\lim _{N \rightarrow \infty} a_{f}(N)=-1$, and also $\lim _{N \rightarrow \infty} a_{f}(N)=1$ if and only if $a_{f}(N)$ is the constant 1 sequence. We will only present here the proof of proposition 4.1.1.

Proof. (Proposition 4.1.1) We prove the proposition by contradiction. Let us assume that there is an $\varepsilon>0$ for which $\forall N$ there exists an $n>N$ for which $\left|g(n)-\inf _{n \in \mathbb{N}^{+}} g(n)\right| \geq \varepsilon$.

Let us consider the value $f_{0}=\inf _{n \in \mathbb{N}^{+}} g(n)+\varepsilon$ and observe the sequence $a_{f_{0}}(N)$. By Proposition 4.1.2, we know that $a_{f_{0}}(N)$ converges. By the assumption, there is a value denoted by $n_{1} \in \mathbb{N}^{+}$at which $g\left(n_{1}\right)>f_{0}$ and also another one denoted by $n_{2} \in \mathbb{N}^{+}$where $g\left(n_{2}\right)<f_{0}$. The latter one means that $a_{f_{0}}\left(n_{2}\right)=-1$, so by the proof of Proposition 4.1.2, $\lim _{N \rightarrow \infty} a_{f_{0}}(N)=-1$, but this contradicts the original assumption.

The value $\inf _{n \in \mathbb{N}^{+}} g(n)$ might be different depending on the $\left(D_{p}, D_{r}, D_{t}\right)$ triplet. However, we can state the following proposition about the maximum of this function.

Proposition 4.1.3. For every $\left(D_{p}, D_{r}, D_{t}\right)$ triplet

$$
\max _{n \in \mathbb{N}^{+}} g(n)=g(1) .
$$

For the proof, see 98]. Using the method from the proof of proposition 4.1.2, we can compute $g(1)$, which will be $g(1)=1.3353 \cdot 10^{-3}$ for every $\left(D_{p}, D_{r}, D_{t}\right)$ value (upon using the same choices for parameters $a, b, c, M$ and $h$ as in the beginning of this chapter).

For the numerical calculations, we will use the same bisection method as in the previous section. However, in this case we fix the value $N$ and search for those values of $f$ where the system (4.6) changes stability. In Figure 4.7, we can see the graph we get for the $\left(D_{p}, D_{r}, D_{t}\right)$ values $(0.1,0.003,0.003)$, where the points above the graph are unstable, and the ones below it are stable. In this way we got the function $g(N)$ with the convergence property proved before.

For $\left(D_{p}, D_{r}, D_{t}\right)$ values $(0.1,0.03,0.03)$ we get the graph in Figure 4.8. Note that we also got the value of $g(1)$ calculated before in both cases.


Figure 4.7: The border of stability (and also the image of the function $g(N)$ ) for $(0.1,0.003,0.003)$. Note that the function $g(N)$ is defined for $N \in \mathbb{N}^{+}$only and the discrete values are connected only for better understanding.

During our examinations, we tested several values, but found only functions either constant or with the property $g(n)=g(2)$ for every even $n$. Because of this, and other observations connected to the proof of proposition 4.1.2, we can state the following conjecture:

Conjecture 4.1.4. For every $\left(D_{p}, D_{r}, D_{t}\right)$ triplet

$$
\min _{n \in \mathbb{N}} g(n)=g(2)
$$

In this way, proposition 4.1.1 has the following form:

$$
g(N) \longrightarrow g(2) \text { as } N \rightarrow \infty
$$

Which also leads to the following property of the $g(n)$ function:

$$
g(2) \leq g(n) \leq g(1)
$$

Without the conjecture, we only have an upper bound for this function, which means that for every $f>g(1)$, the system is unstable.

Therefore, we can conclude that the choice $f=0.001$ in [15] is under the value $g(1)$, so the system can change stability there for some ( $D_{p}, D_{r}, D_{t}$ ) values, but not necessarily (see the ( $0.1,0.03,0.03$ ) case above).


Figure 4.8: The border of stability (and also the image of the $g(N)$ function) for (0.1, 0.03, 0.03).

However, we can make a better choice than $N=10$ using the previous propositions: if our conjecture is true, it is enough to examine our system for $N=2$. If the conjecture is false, then we can also use the $N=2$ value (it is not worse than $N=10$ ), or we have to calculate $\min _{n \in \mathbb{N}} g(n)$ for every triplet $\left(D_{p}, D_{r}, D_{t}\right)$.

### 4.1.3 Numerical solution of the equations

In the previous two sections we examined the way the parameters of diffusion or the $f$ and $N$ values affect the stability of the coexistence equilibrium. However, we do not know what the instability of this point means: what would happen to our system, if the initial conditions were not exactly at the equilibrium, but a small distance from it. Also, it is still an open question whether the stability of the point (4.3) is global or local. It can happen that only the orbits of points in a small neighbourhood of the equilibrium tend to the point (4.3) as $t \longrightarrow \infty$. This would also mean that even if our system was stable at point (4.3), the ecological collapse would still occur on the island if the initial conditions are outside of this neighborhood.

For the investigation of the mentioned surrounding of point (4.3), we use the vector $\mathbf{v} \in \mathbb{R}^{3 N}$ for initial condition, which is defined as

$$
v_{i}:=\left(1-\operatorname{rand}^{\zeta}\right) \frac{1}{N} \frac{M(b-h)}{b+h M f}
$$

where rand is a uniformly distributed random number between 0 and 1 , and $\zeta$ is a fixed number, describing how much we differ from the equilibrium: when $\zeta$ is large we are close to the coexistence equilibrium, and when $\zeta$ is small we are (usually) far from it. (Of course in this case the values of $v_{i}$ will always be between the equilibrium point and the origin - one can also construct a more general initial condition and get similar results.)

For the sake of simplicity, we will use the values $f=0.001$ and $N=10$ in this section. We have seen in Section 4.1.2 that this choice of $f$ is a reasonable one (the system changes stability here). Although we have concluded that $N=2$ is the most costefficient choice, this only applies for the stability of the equilibrium. The larger $N$ values are more interesting ecologically, because we get some data on the geological distribution of our species.

First we examine what the stability of our system means in the point of the actual solutions. For this, we use a fourth order Runge-Kutta method with initial value $\mathbf{v}$. First we test if the point (4.3) is stable globally, or at least locally in a surrounding with considerable sizes. For this, we choose the diffusion parameters to be $(0.1,10,0.1)$ which is a stable point (as it can be seen in Figure 4.5). The results for $\zeta=10$ and $\zeta=2$ are presented in Figure 4.9. As we can see, in both cases the orbits tend to the stable equilibrium point, although this convergence is slower for points further away.

Now we investigate what exactly the instability of point 4.3 means. For this, we choose such parameters for which the equilibrium is unstable, in our case $(0,0.8,0)$. As we can see in Figure 4.10, the ecological catastrophe can happen for various initial conditions.

For the further investigations, we will use another, much more realistic initial condition. Let $\mathbf{w}=\left(w_{i}\right)_{i=1, \ldots, 3 N} \in \mathbb{R}^{3 N}$ be defined as

$$
w_{i}:= \begin{cases}50 & \text { for } i \in\{1,2\} \\ \frac{M}{N} & \text { for } 3 \mid i \\ 0 & \text { otherwise }\end{cases}
$$

In this way, we have $\frac{M}{N}$ trees in every region, while having 50 people and rats in the first sector, which is the moment when the settlers arrived on the island bringing the rats with them. From now on, we will call our system unstable, if there is a region in which one of the species die out, and otherwise stable. Consequently, stability means that our orbit tends to point (4.3) (so it is globally stable), or to any other attracting set. Also,


Figure 4.9: The number of people (blue), rats (red) and trees (orange) in region 1 for diffusion parameters $(0.1,10,0.1)$. On the upper panel we used $\zeta=10$ and on the lower one $\zeta=2$, while on the horizontal axis we have the number of years. The orbit of the rats is very close to the one of the trees', thus it is not visible.
the instability defined this way does not imply the instability of the equilibrium point: it only means that $\mathbf{w}$ is not in the stable neighbourhood of point (4.3).

With the initial vector $\mathbf{w}$, we examine the orbits of solutions in light of the results of Section 4.1.1. For this, we first investigate the effect of the rat diffusion in our system. As we can see it in Figure 4.11, the effect of this parameter only influences the speed of the rats: if $D_{r}=0.01$, then it took the animals two years to get to the other side of the island - however, if $D_{r}=1$, then one year was enough. From the several numerical tests which were made, it can be concluded that $D_{r}$ has a negligible effect on the system's stability, and the other two parameters are the main influential factors.

Now we examine how the change of parameter $D_{p}$ affects our system. We have seen in Section 4.1.1 that for well chosen $\left(D_{r}, D_{t}\right)$ pairs this parameter has a stabilizing effect. Note that this effect only occurred when the system was unstable for small $D_{p}$ values. In


Figure 4.10: The number of people (blue), rats (red) and trees (orange) in the timeintervals $t \in[0,90]$ and $t \in[0,180]$ in region 5 for diffusion parameters $(0,0.8,0)$ and with the number of years on the horizontal axis. As we can see, the collapse can happen for various initial conditions. The orbit of the rats is very close to the one of the trees, thus it is not visible.
our case the small $D_{p}$ values mean that people will not diffuse, so the instability is caused by the rats. However, because of the observations in the previous paragraph, the rats will not have a significant effect on the stability of our system. Therefore, those points which could be stabilized by the $D_{p}$ value are not unstable for small $D_{p}$ values, but will lose their stability for large ones. For the numerical tests, we use the pair $\left(D_{r}, D_{t}\right)=(1,0.01)$. As we can see in Figure 4.12 , when we increase $D_{p}$, our system loses stability.

In Section 4.1.1, we have concluded that the increase of $D_{t}$ makes our system stable for a sufficiently large $D_{t}$ value. For various numerical tests we can see that the increase of $D_{t}$ does not make the unstable systems stable. Also, the increase of this parameter makes the stable systems unstable in a way that first the convergence becomes slower as a periodic orbit appears. For sufficiently large $D_{t}$ values, this orbit vanishes and we


Figure 4.11: The number of people (blue), rats (red) and trees (orange) in region 5 for diffusion parameters $(0.1,0.01,0.1)$ (upper) and ( $0.1,1,0.1$ ) (lower).
get an unstable system (see figure 4.13). In this way we can see that the stability of the coexistence equilibrium does not necessarily correspond to the stability defined for $\mathbf{w}$.

From these results we can see that the stability of the coexistence equilibrium does not necessarily corresponds to the stability of the system. Because of this, further examination of the global behaviour of the system would be beneficial. The results of these studies can be found in Section 4.1.4


Figure 4.12: The number of people (blue), rats (red) and trees (orange) in region 5 for diffusion parameters $(0.02,0.01,0.1)$ (upper) and in region 1 for parameters $(0.02,1,0.1)$ (lower).

### 4.1.4 Global stability of the equilibrium points

As it was mentioned at the end of the previous section, the stability of the coexistence equilibrium point does not necessarily correspond to the stability of the system (which is defined in the previous section). Therefore it is vital to acquire some properties of the behaviour in some other ways. One such way involves the examination of the stability of the other equilibrium points of system (4.2) apart from the previously studied coexistence equilibrium

## The island before the arrival of the settlers

First let us consider the case when there are no people and rats on the island but only trees. This corresponds to the state of the island prior to the arrival of the settlers.


Figure 4.13: The number of people (blue), rats (red) and trees (orange) in region 1 for diffusion parameters $\left(D_{p}, D_{r}\right)=(0.02,0.01)$, and $D_{t}=2$ (upper), $D_{t}=4.2813$ (middle) and $D_{t}=4.2816$ (lower). First our system is stable, then it is still stable but converges slower (there's perhaps a stable periodic orbit around $D_{t}=4.2814$ ), and then it is unstable.

Proposition 4.1.4. If $P^{s}=0$ and $R^{s}=0$ for all $s=1, \ldots, N$ then the system has only one equilibrium point, namely $\left(P^{*}, R^{*}, T^{*}\right)=\left(0,0, \frac{M}{N}\right)$.

Proof. For an equilibrium point we have the equation

$$
0=\frac{d T^{s}}{d t}=b T^{s}\left(1-\frac{T^{s}}{\frac{M}{N}}\right)+D_{t}\left(T^{s-1}-2 T^{s}+T^{s+1}\right)
$$

for every region $s \in\{1,2, \ldots N\}$.
If all the $T^{s}$ values are the same, the diffusion term is zero, and consequently we get the $T^{s}=\frac{M}{N}$ values. If the $T^{s}$ values differ, then (because we have finitely many of them) we have a smallest one - let us denote it by $T^{s_{1}}$. (Note that if there are several smallest values, then we choose the one which has a neighbour with a bigger value.) For this, the diffusion term $T^{s_{1}-1}-2 T^{s_{1}}+T^{s_{1}+1}$ is positive, which means that

$$
b T^{s_{1}}\left(1-\frac{T^{s_{1}}}{\frac{M}{N}}\right)<0
$$

which can only happen if $T^{s_{1}}>\frac{M}{N}$. We will also have a largest value, for which the diffusion term is negative - we will denote it by $T^{s_{2}}$ (also, if there are several largest ones, we choose one which has a neighbour with a smaller value). From this we get that

$$
b T^{s_{2}}\left(1-\frac{T^{s_{2}}}{\frac{M}{N}}\right)>0
$$

which leads to the fact $T^{s_{2}}<\frac{M}{N}$ but this is a contradiction since $\frac{M}{N}<T^{s_{1}}<T^{s_{2}}$ holds.

For the next and later examinations we use the linearized equation, which is

$$
\left(\begin{array}{c}
\frac{d P^{s}}{d t}  \tag{4.10}\\
\frac{d R^{s}}{d t} \\
\frac{d T^{s}}{d t}
\end{array}\right)=S_{s}\left(\begin{array}{c}
P^{s} \\
R^{s} \\
T^{s}
\end{array}\right)+\left(\begin{array}{c}
D_{p}\left(P^{s-1}-2 P^{s}+P^{s+1}\right) \\
D_{r}\left(R^{s-1}-2 R^{s}+R^{s+1}\right) \\
D_{t}\left(T^{s-1}-2 T^{s}+T^{s+1}\right)
\end{array}\right)
$$

where matrix $S_{s}$ is defined as

$$
S_{s}=\left[\begin{array}{ccc}
a-a \frac{2 P^{s}}{T^{s}} & 0 & a \frac{\left(P^{s}\right)^{2}}{\left(T^{s}\right)^{2}} \\
0 & c-c \frac{2 R^{s}}{T^{s}} & c \frac{\left(R^{s}\right)^{2}}{\left(T^{s}\right)^{2}} \\
-h & T^{s}\left(1-\frac{N T^{s}}{M}\right) \frac{-b f N}{\left(1+f N R^{s}\right)^{2}} & \frac{b}{1+f N R^{s}}\left(1-\frac{2 N}{M} T^{s}\right)
\end{array}\right] .
$$

For the equilibrium point $\left(0,0, \frac{M}{N}\right)$ the matrix $S_{s}$ takes the form

$$
S_{s}=\left[\begin{array}{ccc}
a & 0 & 0 \\
0 & c & 0 \\
-h & 0 & -b
\end{array}\right]
$$

Like in the case of the coexistence equilibrium, we apply a Fourier transform to our system, from which we get the equations

$$
\left(\begin{array}{c}
\frac{d x_{q}}{d t} \\
\frac{d y_{q}}{d t} \\
\frac{d z_{q}}{d t}
\end{array}\right)=\left(\begin{array}{ccc}
a-D_{p} C_{q} & 0 & 0 \\
0 & c-D_{r} C_{q} & 0 \\
-h & 0 & -b-D_{t} C_{q}
\end{array}\right)\left(\begin{array}{c}
x_{q} \\
y_{q} \\
z_{q}
\end{array}\right)
$$

in which we used the notation

$$
C_{q}=4 \sin ^{2} \frac{\pi q}{N} .
$$

The three eigenvalues of the matrix are $a-D_{p} C_{q}, c-D_{r} C_{q}$ and $-b-D_{t} C_{q}$, which are all real numbers, so we only have to check whether they are negative or not. It is easy to see that $-b-D_{t} C_{q}$ is always negative but the other two might change their signs. For example, for the first one we need that $a-D_{p} C_{q}<0$ which means $a<D_{p} C_{q}$. However for $q=N$ the right side is zero, but the value of $a$ is positive, which means that the system is always unstable for $q=N$. In this way we can state the following proposition. Proposition 4.1.5. The equilibrium point $\left(0,0, \frac{M}{N}\right)$ is locally unstable, and consequently globally unstable.

This means that with the arrival of the settlers and the rats the system got out of this equilibrium, and can not get back into it, which corresponds to the events on the island.

## The uninhabited island with the rats

Now let us examine the case of $P^{s}=0(s=1, \ldots, N)$. This corresponds to the hypothetical situation when there are no people on the island, but the rats live there. Although in the theories of Hunt the rats are brought by the settlers, our dynamical system will have an equilibrium with such properties, so it is beneficial to consider the stability of this point also.

Proposition 4.1.6. If $P^{s}=0(s=1, \ldots, N)$, the system has only one equilibrium point, namely $\left(P^{*}, R^{*}, T^{*}\right)=\left(0, \frac{M}{N}, \frac{M}{N}\right)$.

Proof. Note that the division of $b$ by $1+f N R^{s}$ does not change the sign of the term, which means that the same method can be applied here as in the proof of proposition 4.1.4.

The matrix at this point is in the form

$$
S_{s}=\left[\begin{array}{ccc}
a & 0 & 0 \\
0 & -c & c \\
-h & 0 & \frac{-b}{1+f M}
\end{array}\right]
$$

After the application of the Fourier transform we get the equation

$$
\left(\begin{array}{c}
\frac{d x_{q}}{d t} \\
\frac{d y_{q}}{d t} \\
\frac{d z_{q}}{d t}
\end{array}\right)=\left(\begin{array}{ccc}
a-D_{p} C_{q} & 0 & 0 \\
0 & -c-D_{r} C_{q} & c \\
-h & 0 & \frac{-b}{1+f M}-D_{t} C_{q}
\end{array}\right)\left(\begin{array}{c}
x_{q} \\
y_{q} \\
z_{q}
\end{array}\right)
$$

The three eigenvalues of the matrix are $a-D_{p} C_{q},-c-D_{r} C_{q}$ and $\frac{-b}{1+f M}-D_{t} C_{q}$, which are real values. The second and third ones are always negative, but the first is positive for $q=N$, which means that the matrix is unstable.

Proposition 4.1.7. The equilibrium point $\left(0,0, \frac{M}{N}\right)$ is locally unstable, and consequently globally unstable.

This means that the rats and the trees cannot survive the extinction of people on the island.

## The equilibrium of the rat-free model

Now let us consider the case of $R^{s}=0(s=1, \ldots, N)$, which corresponds to the original, rat-free model. In this case the matrix we have takes the form

$$
S_{s}=\left[\begin{array}{ccc}
-a & 0 & a \\
0 & c & 0 \\
-h & \frac{-f M h(b-h)}{b} & 2 h-b
\end{array}\right]
$$

After the Fourier transformation we get the system

$$
\left(\begin{array}{c}
\frac{d x_{q}}{d t} \\
\frac{d y_{q}}{d t} \\
\frac{d z_{q}}{d t}
\end{array}\right)=\left(\begin{array}{ccc}
-a-D_{p} C_{q} & 0 & a \\
0 & c-D_{r} C_{q} & 0 \\
-h & \frac{-f M h(b-h)}{b} & 2 h-b-D_{t} C_{q}
\end{array}\right)\left(\begin{array}{c}
x_{q} \\
y_{q} \\
z_{q}
\end{array}\right)
$$

One of the eigenvalues of the stability matrix is $c-D_{r} C_{q}$, which is positive for $q=N$. We did not answer the question whether this equilibrium point is unique or not. However, it is easy to see that the second row of the stability matrix does not depend on which equilibrium point we study in our system (since $R^{s}=0$ ). Thus, every such matrix will have a positive eigenvalue for $q=N$.

Proposition 4.1.8. Every equilibrium point for $R^{s}=0(s=1, \ldots, N)$ is locally unstable and consequently globally unstable.

This also means that the rats cannot die out without the extinction of the other two species on the island. It is an important remark that although the Polynesian rat had disappeared from Easter island, it was not because of the people, but because of the European rats brought to the island by the settlers in the eighteenth century - of course, this latter phenomenon is not described by our model.

## The state of extinction

Now let us consider the phenomena that occur near the origin in our system. In [14], Basener et al. only state that this equilibrium is a singular one, therefore cannot be examined with the previous methods. However, one of the most important states (next to the coexistence equilibrium) is extinction, which means that the study of this point cannot be neglected.

It is also not clear whether the origin is an equilibrium point: if we take the limit $\lim _{\mathbf{Q} \longrightarrow \mathbf{0}} \frac{d P^{s}}{d t}\left(\right.$ or $\left.\lim _{\mathbf{Q} \longrightarrow \mathbf{0}} \frac{d R^{s}}{d t}\right)$ where $\mathbf{Q}$ is the vector of all the $P^{s}, R^{s}$ and $T^{s}$ values (and here $\mathbf{0}$ means the all zero vector of the same size), we get a term in the form $\frac{0}{0}$. Also, it can be shown that the limit $\lim _{\mathbf{Q} \longrightarrow \mathbf{0}} \frac{d P^{s}}{d t}\left(\right.$ or $\left.\lim _{\mathbf{Q} \longrightarrow \mathbf{0}} \frac{d R^{s}}{d t}\right)$ does not exist: take e.g. the sequences $\left(P_{k}^{s}, T_{k}^{s}\right)=\left(\frac{1}{k}, \frac{1}{k}\right)$ and $\left(P_{\ell}^{s}, T_{\ell}^{s}\right)=\left(\frac{1}{\ell}, \frac{1}{\ell^{2}}\right)$ and compute the corresponding values of $\frac{d P_{k}^{s}}{d t}$ and $\frac{d P_{\ell}^{s}}{d t}$. Then,

$$
\lim _{k \rightarrow \infty} \frac{d P_{k}^{s}}{d t} \neq \lim _{\ell \rightarrow \infty} \frac{d P_{\ell}^{s}}{d t}
$$

However, our original equations aren't even defined in the case of $T^{s}=0$ : because of this, we are going to modify our original system of equations (4.2) in the following way.

Let us say that equations (4.2) only hold when $P^{s}, R^{s}, T^{s}>\varepsilon$ for some $\varepsilon>0$ sufficiently small constant (e.g. $\varepsilon<1$ ). This can be assumed since the processes from which the equations are derived only take place when the amount of the species is not zero, e.g. $P^{s}>1$ (this is also a reasonable assumption since $P^{s}=1$ means one person). Because of this, one can extend the system of equations (4.2) in a way that (4.2) holds only when $P^{s}, R^{s}, T^{s}>\varepsilon$ and

$$
\left\{\begin{array}{l}
P^{s}=0 \\
R^{s}=0 \\
T^{s}=0
\end{array}\right.
$$

otherwise.
It is evident that in this case our solution is not differentiable when $P^{s}=\varepsilon, R^{s}=\varepsilon$ or $T^{s}=\varepsilon$ and it is not even continuous there. However, since the case $P^{s} \leq 1$ means that the reproduction of people is impossible on the island, then it is reasonable to consider the state $P^{s}=1$ to be the extinction of humans. Similar arguments hold for the other species
also, possibly with other constants (when they are unable to reproduce, e.g. $R^{s} \leq 1$ or $T^{s}<1$ ). By these assumptions, the state of extinction, e.g. when $P^{s}, R^{s}, T^{s}=0$ for all $s=1, \ldots, N$ can be thought of as an equilibrium point (since the derivatives are zero there).

Now let us discuss the stability of this equilibrium point. By the construction detailed above, we cannot use the methods which were applied previously in the case of the other points. Instead of this, we can use a geometric method discussed in [100].

Proposition 4.1.9. There exists a neighbourhood around the origin in which $\frac{d T^{s}}{d t}>0$.
Seemingly this result means that the origin is unstable, therefore extinction cannot happen on the island, and it would also contradict the graphs we got in Section 4.1.3. However, by the proof it can also be seen that as we approach the origin, this positive regions decreases until the point when the whole $\mathbf{Q}>\mathbf{0}$ region (the inequality is meant element-wise) has a negative $\frac{d T^{s}}{d t}$ value. The reason for this is that the size of the aforementioned region depends on the values of $T$ in the neighboring regions, so if these values decrease, this region will also shrink. It is also clear from the proof that in this case this region has negative $\frac{d P^{s}}{d t}$ and $\frac{d R^{s}}{d t}$ values too.

In this way we can state that although there are some values from which there is no orbit tending towards the origin, but as we approach zero from outside of this region, the area of these values decreases. Consequently we can conclude that orbits approaching the origin from a suitable direction might tend to it while others will not. In our numerical solutions we found orbits which had the convergence property (see Figure 4.10) and also some which do not (see orbits on the third graph of Figure 4.13. first they turn back, but then they get to the origin).

### 4.2 Continuous model

In this section we consider a continuous version of the previous, discrete model, meaning that instead of splitting the coast of the island into $N$ regions, we model it as a connected one dimensional space. Then we observe the effect of the increase of the diffusion of the trees, and get similar results to the ones proved in the previous case.

### 4.2.1 Construction of the model

Let us suppose that we have a partition of the coast of the island into $N$ regions. Because of the assumption that there are no differences in the densities of the species radius-wise (meaning that the number of each of the species is the same along a line drawn from the center of the island towards the coast), we can think of the coast as a one-dimensional line and model our problem as a one-spatial-dimensional one with periodic boundary conditions. In this way our domain will be an interval, let us denote it by $\Omega=[0, \mathcal{L})$ and let its partition be $\left\{\Omega^{\alpha}\right\}_{\alpha=1,2, \ldots N}$. Let $P(t, x), R(t, x)$ and $T(t, x)$ be the density of people, rats and trees at point $x \in \Omega$ at time $t$, respectively. Thus, the number of individuals in region $\alpha$ at time $t$ denoted by $Q^{\alpha}(t)$ can be calculated as

$$
Q^{\alpha}(t)=\int_{\Omega^{\alpha}} Q(t, x) \mathrm{d} x
$$

where $Q \in\{P, R, T\}$, and $\Omega^{\alpha}$ denotes the domain corresponding to the region indexed by $\alpha$. Now let us assume that the density of a given species is the same inside a given region. (For a sufficiently large $N$, this is a good approximation.) In this way we get the equation

$$
\mathcal{L}^{\alpha} Q(t, x)=Q^{\alpha}(t)
$$

where $\mathcal{L}^{\alpha}$ is the length of region $\Omega^{\alpha}$ and $x \in \Omega^{\alpha}$ arbitrary. For the sake of simplicity, let us suppose that the lengths of the regions are equal. Writing these into equation (4.2) omitting the diffusion terms, we get the following:

$$
\begin{align*}
& \frac{\partial P(t, x)}{\partial t}=a P(t, x)\left(1-\frac{P(t, x)}{T(t, x)}\right) \\
& \frac{\partial R(t, x)}{\partial t}=c R(t, x)\left(1-\frac{R(t, x)}{T(t, x)}\right)  \tag{4.11}\\
& \frac{\partial T(t, x)}{\partial t}=\frac{b}{1+f \mathcal{L} R(t, x)} T(t, x)\left(1-\frac{T(t, x)}{\frac{M}{\mathcal{L}}}\right)-h P(t, x),
\end{align*}
$$

where we used that $N \mathcal{L}^{\alpha}=\mathcal{L}$. Let us introduce the notations $\widetilde{f}=f \mathcal{L}$ and $\widetilde{M}=M / \mathcal{L}$. In order to model the motion of the species we add some diffusion terms to the equations:

$$
\begin{align*}
& \frac{\partial P(t, x)}{\partial t}=a P(t, x)\left(1-\frac{P(t, x)}{T(t, x)}\right)+D_{P} \Delta P(t, x) \\
& \frac{\partial R(t, x)}{\partial t}=c R(t, x)\left(1-\frac{R(t, x)}{T(t, x)}\right)+D_{R} \Delta R(t, x)  \tag{4.12}\\
& \frac{\partial T(t, x)}{\partial t}=\frac{b}{1+\widetilde{f} R(t, x)} T(t, x)\left(1-\frac{T(t, x)}{\widetilde{M}}\right)-h P(t, x)+D_{T} \Delta T(t, x)
\end{align*}
$$

in which $\Delta$ is the Laplace operator taken in the second variable.
As mentioned before, we will assign a periodic boundary condition to the problem, meaning that:

$$
\partial_{x}^{\beta} Q(t, 0)=\partial_{x}^{\beta} Q(t, \mathcal{L}), \quad \beta \in\{0,1,2\},
$$

where $Q \in\{P, R, T\}$, and $\partial_{x}$ denotes partial derivative with respect to the second coordinate, in the sense that at point $x=0$ it is a right derivative, and at point $x=\mathcal{L}$ a left derivative. The previous system can be rewritten as follows:

$$
\left\{\begin{align*}
\partial_{t} \mathcal{Q}(t, x) & =F(\mathcal{Q}(t, x))+D \Delta \mathcal{Q}(t, x)  \tag{4.13}\\
\partial_{x}^{\beta} \mathcal{Q}(t, 0) & =\partial_{x}^{\beta} \mathcal{Q}(t, \mathcal{L}), \quad \beta \in\{0,1,2\} \\
\mathcal{Q}(0, x) & =\varphi(x)
\end{align*}\right.
$$

where $\mathcal{Q}(t, x)=(P(t, x), R(t, x), T(t, x)), F$ is the $\mathbb{R}^{3} \rightarrow \mathbb{R}^{3}$ function describing the interactions between the species, $\partial_{t}$ is the time derivative, $\partial_{x}$ is meant as before but element-wise, $D$ denotes the diagonal matrix containing the diffusion coefficients and $\varphi(x)$ gives the initial distribution of the species on the island. In this way we get a reaction-diffusion equation with periodic boundary conditions. We are searching for the classical solution $Q(t, x)$ which is in $C^{1,2}(\Omega)$ i.e. it is continuously differentiable in its time variable once, and twice in the spatial coordinate.

### 4.2.2 The effect of the diffusion of the trees

A convenient way to examine the behavior of system (4.12) is to search for its constant stationary solutions and check their stability properties, i.e. those solutions which take the same value at every point of the shore. It is easy to see that this is a bi-
jection between these and the equilibrium points of the system (4.2). In other words, $Q(t, x)=\left(P^{*}, R^{*}, T^{*}\right)$ is a stationary solution of 4.12) if and only if $Q(t)=\left(\mathcal{L} P^{*}, \mathcal{L} R^{*}, \mathcal{L} T^{*}\right)$ is an equilibrium point of 4.2), where $P^{*}, R^{*}$ and $T^{*}$ are constants. Because of this, system (4.12) has only one coexistence constant stationary solution, which is

$$
\begin{equation*}
P^{*}(t, x)=R^{*}(t, x)=T^{*}(t, x) \equiv \frac{\widetilde{M}(b-h)}{b+h \widetilde{M} \widetilde{f}}, \quad \forall x \in \Omega, t \in \mathbb{R}^{+} \tag{4.14}
\end{equation*}
$$

This is a corollary of the similar property of (4.2), which can be easily proved.
For the examination of the stability of this solution, we are going to use the following method (also described e.g. in [84]). Let us linearize around the aforementioned stationary solution, and search for perturbations in the form

$$
\begin{equation*}
w(t, x)=w_{0} e^{\lambda t} W(x) \tag{4.15}
\end{equation*}
$$

Because we have periodic boundary conditions, any solutions can be written as the sum of sinusoidal waves of the form

$$
W(x)=c_{1} e^{i k x}+c_{2} e^{-i k x}
$$

Our steady solution is stable if all of the growth rates $\lambda$ have negative real parts, and is unstable otherwise.

If we substitute (4.15) into system (4.12), after a brief calculation we get the following system of equations:

$$
\begin{aligned}
& \lambda P(t)=-k^{2} D_{P} P(t)-a P(t)+a T(t) \\
& \lambda R(t)=-k^{2} D_{R} R(t)-c R(t)+c T(t) \\
& \lambda T(t)=-k^{2} D_{T} T(t)-h P(t)+A R(t)+B T(t)
\end{aligned}
$$

in which we used the notations $\widetilde{A}=\frac{-\widetilde{f} \widetilde{M} h(b-h)}{b(1+\widetilde{f} \widetilde{M})}$ and $\widetilde{B}=\frac{\widetilde{f} \widetilde{M} h-b+2 h}{1+\widetilde{f} \widetilde{M}}$. This system
can be rewritten as a matrix in the following way

$$
\left[\begin{array}{ccc}
-\lambda-k^{2} D_{P}-a & 0 & a  \tag{4.16}\\
0 & -\lambda-k^{2} D_{R}-c & c \\
-h & \widetilde{A} & -\lambda-k^{2} D_{T}+\widetilde{B}
\end{array}\right]\left(\begin{array}{c}
P(t) \\
R(t) \\
T(t)
\end{array}\right)=\left(\begin{array}{l}
0 \\
0 \\
0
\end{array}\right)
$$

Our system is stable if $\lambda$ has negative real part, meaning that the matrix

$$
\mathcal{M}:=\left[\begin{array}{ccc}
-k^{2} D_{P}-a & 0 & a  \tag{4.17}\\
0 & -k^{2} D_{R}-c & c \\
-h & \widetilde{A} & -k^{2} D_{T}+\widetilde{B}
\end{array}\right]
$$

has negative eigenvalues.
The next key observation is that this matrix $\mathcal{M}$ is very similar to the one which was observed in Section 4.1.1 (and also in [99]). The only difference between the two is that in that case the matrix had Neumann eigenvalues instead of the terms $k^{2}$, but because both are positive, they behave the same way. Consequently, the proofs presented there (using the Routh-Hurwitz criteria) are applicable also for this problem, and thus, we can state similar theorems.

Theorem 4.2.1. Let us suppose that the model parameters satisfy the condition

$$
\begin{equation*}
\widetilde{B}^{2}-\widetilde{A} c-a h<0 . \tag{4.18}
\end{equation*}
$$

Moreover, let us suppose that $D_{P}$ and $D_{R}$ are fixed positive diffusion values. Then, if system (4.12) is stable for $D_{T}=0$ then it is stable for all positive $D_{T}$ values.

Remark 4.2.1. The values of the parameters originally chosen in 13,15 are the following. ( $\mathcal{L}$ is chosen to be 60000 , since the coast of the island is approximately 60000 meters long, however, it does not affect the condition.)

$$
a=0.03, \quad b=1, \quad c=10, \quad M=12000, \quad f=0.001, \quad h=0.25, \quad \mathcal{L}=60000 .
$$

Condition 4.18 is satisfied if $f>8.6751 \times 10^{-6}$ (or $\widetilde{f}>0.52051$ ), which can be assumed since rats have a significant effect on the reproduction rate of the trees.

In this way, if the system was stable in the case of no tree diffusion, it will remain so even if $D_{T}$ is increased, so the region of stability cannot shrink as the diffusion is increased.

Theorem 4.2.2. Let $D_{P}$ and $D_{R}$ be two fixed positive numbers, and suppose that the following conditions hold:

$$
\begin{align*}
\widetilde{B} & <\min \{a+c, \widetilde{A}+h\}, \\
\widetilde{B}\left(\widetilde{A} c+a h+(a+c)^{2}\right) & <c^{2}(\widetilde{A}+a)+\widetilde{B}^{2}(a+c)+a^{2}(c+h) . \tag{4.19}
\end{align*}
$$

Then there is a positive number $\bar{D}_{T}$ such that the system (4.12) is stable for all $D_{T}>\bar{D}_{T}$ values.

Remark 4.2.2. Condition (4.19) is satisfied with the parameter values described above.
Consequently, the increase of $D_{T}$ makes our system stable even if it was unstable for small values of the diffusion of the trees.

Now let us examine whether the previous analytic results can be confirmed by some numerical experiments. For these, a finite difference scheme is used in the following way. Let us consider a grid $\mathcal{G}$ on our interval $[0, \mathcal{L})$ which consists of $\mathcal{N}$ number of points $\left\{\omega_{i}: \frac{\mathcal{L}}{\mathcal{N}}\right\}_{i=0}^{\mathcal{N}-1}$, and let us denote the numerical approximation of our function $\mathcal{Q}(t, x)$ at these points by $\mathcal{Q}_{i}(t)$. Also, let us approximate the diffusion terms $D \Delta \mathcal{Q}(t, x)$ by $D(\mathcal{N} / \mathcal{L})^{2}\left(\mathcal{Q}_{i-1}(t)-2 \mathcal{Q}_{i}(t)+\mathcal{Q}_{i+1}(t)\right)$. Note that this way we got the same system as (4.2), meaning that the numerical solution of this semi-discretized problem would be the same as the one already calculated in Section 4.1.3. In that a fourth order Runge-Kutta method is used with initial vector $\mathbf{w} \in \mathbb{R}^{3 \mathcal{N}}$ which has elements defined as

$$
w_{i}:= \begin{cases}0.001 & \text { for } i \in\left\{3 j-2,3 j-1 \mid j \in \mathbf{Z}^{+}, j \leq 0.1 \mathcal{N}\right\} \\ \widetilde{M} & \text { for } 3 \mid i \\ 0 & \text { otherwise }\end{cases}
$$

In this way, we have trees with density of $\widetilde{M}$ in every region, and a relatively small density of people and rats in a small part of our coast, which corresponds to the moment when the settlers arrived on the island bringing the rats with them.

With the initial vector $\mathbf{w}$ we examine the way the previous analytic results occur in the terms of the orbits of solutions. On the left panel of Figure 4.14 we can see that for the
diffusion parameters $\left(D_{P}, D_{R}, D_{T}\right)=(0.02,0.01,0.0005)$ the system is unstable, meaning that the populations die out. However, if the diffusion of the trees is increased and the parameters $(0.02,0.01,0.001)$ are considered, then the solutions tend to the stationary solution (4.14), which confirms the analytic results.



Figure 4.14: The number of people (blue), rats (red) and trees (orange) at point $x=0$ with $\mathcal{N}=100$ spatial mesh points for diffusion parameters ( $0.02,0.01,0.0005$ ) (left) and for parameters $(0.02,0.01,0.001)$ (right). As we can see, the increase of the diffusion of the trees stabilizes the stationary solution corresponding to coexistence. Note that the density of rats is very close to the density of the trees, thus it is not visible.

### 4.3 Conclusions

In the previous pages we extended the model describing the dynamics of Easter Island originally proposed by Basener et al. In Section 4.1.1 we studied the way the addition of the parameter $D_{t}$ affects the stability of our system at the coexistence equilibrium. We have seen that while $D_{t}$ has a stabilizing effect on point (5.5) and $D_{r}$ has the opposite, $D_{p}$ can have a stabilizing one for small, and destabilizing one for larger values. In Section 4.1 .2 we have also examined the best choice of the parameters $f$ and $N$ for interesting results and time-efficient computing.

After these, we investigated the numerical solutions of the original system and the meaning of the stability of point (5.5) in the terms of the orbits of the points near it. We have found that both stability and instability have the properties expected before. We also modelled the scenario which could have happened on the island. We found out that for some parameter values a stable equilibrium can occur, but the ecological catastrophe is also possible.

As mentioned before, when the first European settlers arrived on the island, they
found no trees and very few rats on the island. As we can see in Figures 4.10 and 4.11 , the number of people is always larger at the time of the catastrophe than the number of trees and rats. Because of this, we can suspect that after the trees (and rats) disappeared from our island, the inhabitants looked for other resources on the island, for example tried to fish from the coast.

One of the remarkable innovations in the theory of Hunt was the fact that the ecological collapse happened much faster than expected. In our numerical test we have seen that in most cases the collapse happens about 180 years after the arrival of the settlers (see the right side of Figure 4.12). However, with a careful choice of parameters we can extend this moment for a few hundred years, and can even have a periodic solution (see Figure 4.13), although these happen only for a carefully chosen triplets of diffusion, which is highly unlikely to happen in real life.

We have also seen that sometimes the numerical results did not correspond to the local stability of the coexistence equilibrium point. Because of this, we studied the stability of the other equilibria of the system in Section 4.1.4, where we found that the only other one with a stable region is the origin, which corresponds to the original ecological model.

In Section 4.2, the previous model was modified in a way that the previous system of ordinary differential equations was transformed into a system of partial differential equations. It turned out that the stabilizing effect of the diffusion of the trees holds in this case too.

## Chapter 5

## Two-dimensional model

In the previous chapter we proposed a model in which we looked at Easter island as an island which has a big volcano in the middle and only its shore is habitable. However, Easter island has a shape of an isosceles right triangle with hills near the three vertices (see Figure 5.1). Therefore, it is more accurate to propose a two-dimensional model. On the following pages we show that similar properties hold in this two-dimensional case as in the previous one, namely that the increase of the tree diffusion stabilizes the system.


Figure 5.1: The Island of Rapa Nui with the two volcanoes.
A natural extension of equation (4.2) would be to construct a grid on the island (see Figure 5.2. However, note that in the one dimensional case we used Fourier transformation which required that regions 1 and $N$ are connected in that model. It is easy to see that in two dimensions we do not have such properties, so the use of the aforementioned transformation is not beneficial. Also, such two-dimensional grids are much more complicated to handle in the stability investigations than the one dimensional ones. Because of all of these, we will not use such models, but will construct a system of partial differential equations to describe not only the evolution in time, but also the spatial propagation of the different species.


Figure 5.2: A possible grid for the island.

### 5.1 Construction of the model

Although we will not use the grid method mentioned above, it is easier to understand the continuous model if we construct it from such a semi-discrete one. First let us neglect the diffusion between the regions of our two-dimensional one and assume that the same processes take place inside every region as in the one-dimensional case. Then, this twodimensional version of (4.2) takes the form

$$
\begin{align*}
& \frac{d P^{\alpha}(t)}{d t}=a P^{\alpha}(t)\left(1-\frac{P^{\alpha}(t)}{T^{\alpha}(t)}\right) \\
& \frac{d R^{\alpha}(t)}{d t}=c R^{\alpha}(t)\left(1-\frac{R^{\alpha}(t)}{T^{\alpha}(t)}\right)  \tag{5.1}\\
& \frac{d T^{\alpha}(t)}{d t}=\frac{b}{1+f \widetilde{N} R^{\alpha}(t)} T^{\alpha}(t)\left(1-\frac{T^{\alpha}(t)}{\frac{M}{\widetilde{N}}}\right)-h P^{\alpha}(t),
\end{align*}
$$

where $\alpha$ denotes the index of the examined region, $\widetilde{N}$ is the number of regions and $Q^{\alpha}(t)$ $(Q \in\{P, R, T\})$ gives the number of the given species in region $\alpha$ at a given time $t$.

Let us represent the island as a domain on $\mathbb{R}^{2}$ denoted by $\Omega$, and let its boundary be $\partial \Omega$. Also, let us examine the number of individuals of a given species in a given region on the island. If we denote by $P(t, x), R(t, x)$ and $T(t, x)$ the density of the given species at time $t$ at a point $x$ on the island, then the number of individuals of a given species can be calculated in the following way:

$$
\int_{\Omega^{\alpha}} Q(t, x) \mathrm{d} x=Q^{\alpha}(t),
$$

where $Q \in\{P, R, T\}$ and $\Omega^{\alpha}$ denotes the domain corresponding to the region indexed by $\alpha$. Now let us assume that the density of a given species is the same inside a given
region. (For a sufficiently large $\widetilde{N}$, this is a good approximation.) Consequently, we get the equation

$$
\mathcal{A}^{\alpha} Q(t, x)=Q^{\alpha}(t),
$$

where $\mathcal{A}^{\alpha}$ is the area of region $\Omega^{\alpha}$ and $x \in \Omega^{\alpha}$ arbitrary. For the sake of simplicity let us suppose that the areas of the regions are equal. Writing these into equation (5.1) we get the following:

$$
\begin{align*}
& \frac{\partial P(t, x)}{\partial t}=a P(t, x)\left(1-\frac{P(t, x)}{T(t, x)}\right) \\
& \frac{\partial R(t, x)}{\partial t}=c R(t, x)\left(1-\frac{R(t, x)}{T(t, x)}\right)  \tag{5.2}\\
& \frac{\partial T(t, x)}{\partial t}=\frac{b}{1+f \mathcal{A} R(t, x)} T(t, x)\left(1-\frac{T(t, x)}{\frac{M}{\mathcal{A}}}\right)-h P(t, x),
\end{align*}
$$

where we used that $\widetilde{N} \mathcal{A}^{\alpha}=\mathcal{A}$, where $\mathcal{A}$ denotes the area of the island. Let us introduce the notations $\widehat{f}=f \mathcal{A}$ and $\widehat{M}=M / \mathcal{A}$. In order to model the motion of the species we add some diffusion terms to the equations:

$$
\begin{align*}
& \frac{\partial P(t, x)}{\partial t}=a P(t, x)\left(1-\frac{P(t, x)}{T(t, x)}\right)+D_{P} \Delta P(t, x) \\
& \frac{\partial R(t, x)}{\partial t}=c R(t, x)\left(1-\frac{R(t, x)}{T(t, x)}\right)+D_{R} \Delta R(t, x)  \tag{5.3}\\
& \frac{\partial T(t, x)}{\partial t}=\frac{b}{1+\widehat{f} R(t, x)} T(t, x)\left(1-\frac{T(t, x)}{\widehat{M}}\right)-h P(t, x)+D_{T} \Delta T(t, x) .
\end{align*}
$$

We also know the boundary conditions for this problem, because the species will not leave the island, so we can state a homogeneous Neumann boundary condition:

$$
\partial_{\mathbf{v}} P(t, x)=\partial_{\mathbf{v}} R(t, x)=\partial_{\mathbf{v}} T(t, x)=0, \quad x \in \partial \Omega
$$

where $\partial_{\mathbf{v}}$ denotes the directional derivative taken on the vector $\mathbf{v}$ which is the outward normal vector of the set $\Omega$. The previous system can be rewritten as follows:

$$
\left\{\begin{array}{rl}
\partial_{t} U(t, x) & =F(U(t, x))+D_{U} \Delta U(t, x)  \tag{5.4}\\
\partial_{\mathbf{v}} U(t, x) & =0, \\
U(0, x) & =\varphi(x)
\end{array} \quad \forall x \in \partial \Omega,\right.
$$

where $U(t, x)=(P(t, x), R(t, x), T(t, x)), F$ is the $\mathbb{R}^{3} \rightarrow \mathbb{R}^{3}$ function describing the interactions between the species, $\partial_{t}$ is the time derivative, $D_{U}$ denotes the diagonal matrix containing the diffusion coefficients and $\varphi(x)$ is a function describing the initial distribution of the species on the island. Therefore, we got a reaction-diffusion equation with Neumann boundary conditions. We are searching for the classical solution $U(t, x)$ which is in $C^{1,2}(\Omega)$, i.e. it is continuously differentiable in its time variable once, and in the spatial coordinate twice.

### 5.2 The effect of the diffusion of trees

To understand the behaviour of system (5.4), we will examine the stability of its constant stationary solutions, i.e. which take the same value at every point of the island and do not change in time. It is easy to see that these are equivalent to the equilibrium points of the system (4.1). In other words, $U(t, x)=\left(P^{*}, R^{*}, T^{*}\right)$ is a stationary solution of (5.4) if and only if $U(t)=\left(\mathcal{A} P^{*}, \mathcal{A} R^{*}, \mathcal{A} T^{*}\right)$ is an equilibrium point of (4.1), where $P^{*}, R^{*}$ and $T^{*}$ are constants. Because of this, we can state the following lemma:

Lemma 5.2.1. System (5.4) has only one coexistence constant stationary solution, which is

$$
\begin{equation*}
P^{*}(t, x)=R^{*}(t, x)=T^{*}(t, x) \equiv \frac{\widehat{M}(b-h)}{b+h \widehat{M} \widehat{f}}, \quad \forall x \in \Omega, t \in \mathbb{R}^{+} \tag{5.5}
\end{equation*}
$$

Proof. This is a corollary of the similar property of (4.1), which can be easily proved.
Now we would like to state similar theorems about the stability properties to the ones given in the one dimensional case in Chapter 4. For this, we will use the following result.

Theorem 5.2.1 (See [25, 83). Consider the following reaction-diffusion equation given on a bounded domain $\Omega$ with smooth boundary in a finite dimensional Euclidean space:

$$
\left\{\begin{array}{lr}
\partial_{t} u(t, x)=F(u(t, x))+D \Delta u(t, x), & \forall x \in \Omega,  \tag{5.6}\\
\partial_{\mathbf{v}} u(t, x)=0, & \forall x \in \partial \Omega
\end{array}\right.
$$

in which $D$ is a positive valued diagonal matrix. Let us suppose that it has a constant stationary solution, namely $u(t, x)=C^{*}$. This is asymptotically stable if and only if the real parts of the eigenvalues of the matrices in the form $L-\lambda_{n} D$ are negative for every
$n \in \mathbb{N}$, in which $L$ is the linearization of $F(u(t, x))$ around the point $C^{*}$, and $\lambda_{n}$ is a Neumann eigenvalue, i.e. it is the solution of the eigenvalue problem:

$$
\left\{\begin{align*}
-\Delta w_{n}(x) & =\lambda_{n} w_{n}(x), & & \forall x \in \Omega  \tag{5.7}\\
\partial_{\mathbf{v}} w_{n}(x) & =0, & & \forall x \in \partial \Omega
\end{align*}\right.
$$

It is well known that on a domain described in the theorem the eigenvalues of (5.7) are $0=\lambda_{0} \leq \lambda_{1} \leq \lambda_{2} \leq \ldots$ (see e.g. [30]).

Now we have to calculate the matrix $L-\lambda_{n} D$ around equilibrium (5.5). Because of the aforementioned connection between systems (4.1) and (5.4), $L$ is similar to the matrix in equation (4.4) in Section 4.1. namely

$$
L=\left[\begin{array}{ccc}
-a & 0 & a \\
0 & -c & c \\
-h & \frac{-\widehat{f} \widehat{M} h(b-h)}{b(1+\widehat{f M})} & \frac{\widehat{f M} h-b+2 h}{1+\widehat{f M}}
\end{array}\right]
$$

From now on, let us use the notations $\widehat{A}=\frac{-\widehat{f M} h(b-h)}{b(1+\widehat{f M})}$ and $\widehat{B}=\frac{\widehat{f M} h-b+2 h}{1+\widehat{f M}}$. Hence, the matrix $L-\lambda_{n} D$ takes the form

$$
L-\lambda_{n} D=\left[\begin{array}{ccc}
-a-\lambda_{n} D_{P} & 0 & a  \tag{5.8}\\
0 & -c-\lambda_{n} D_{R} & c \\
-h & \widehat{A} & \widehat{B}-\lambda_{n} D_{T}
\end{array}\right]
$$

Now we have to determine whether the eigenvalues of this matrix have a negative real part. For this, we use the Routh-Hurwitz criteria [63, 88] which was already stated in Theorem 4.1.1. We will use these three conditions to determine the signs of the eigenvalues of matrix $L-\lambda_{n} D$.

First we would like to show that the increase of the diffusion term in the third equation of (5.3) does not destroy stability, i.e. if the constant solution was stable for $D_{T}=0$, then it will remain stable for any positive value of $D_{T}$. For this, we will use the following theorem, which is an extension of the previous one, this time allowing zero values in $D$.

Theorem 5.2.2 (Corollary 2 in [25). Consider the following reaction-diffusion equation given on a bounded domain $\Omega$ with smooth boundary in a finite dimensional Euclidean space:

$$
\begin{cases}\partial_{t} u(t, x)=F(u(t, x))+D \Delta u(t, x), & \forall x \in \Omega,  \tag{5.9}\\ \partial_{\mathbf{v}} u(t, x)=0, & \forall x \in \partial \Omega,\end{cases}
$$

in which the diagonal matrix $D$ has the form

$$
D=\left(\begin{array}{cc}
\widetilde{D} & 0 \\
0 & 0
\end{array}\right) .
$$

Let us suppose that (5.6) has a constant stationary solution, namely $u(t, x)=C^{*}$. Let matrix $L$ be the linearization of $F(u(t, x))$ around the point $C^{*}$ and let us write matrix $L$ in the form

$$
L=\left(\begin{array}{cc}
\widetilde{L} & L_{1} \\
L_{2} & L_{3}
\end{array}\right)
$$

where $\widetilde{L}$ has the same size as $\widetilde{D}$. We also suppose that all the eigenvalues of $L_{3}$ have a negative real part.

The constant solution $u(t, x)=C^{*}$ is asymptotically stable if and only if the real parts of the matrices $L-\lambda_{n} D$ are negative ( $n \in \mathbb{N}$ ), in which $\lambda_{n}$ is a Neumann eigenvalue defined as the solution of (5.7).

We can now state the result regarding the stability of our system. From now on we call our system (5.4) stable if the coexistence constant solution (5.5) is stable, and call it unstable if (5.5) is unstable.

Theorem 5.2.3. Let us suppose that the model parameters satisfy the conditions

$$
\begin{gather*}
\widehat{B}^{2}-\widehat{A} c-a h<0, \\
\widehat{f}<\frac{b-2 h}{\widehat{M} h} . \tag{5.10}
\end{gather*}
$$

Moreover, let us suppose that $D_{P}$ and $D_{R}$ are fixed positive diffusion values. Then, if system (5.4) is stable for $D_{T}=0$ then it is stable for all positive $D_{T}$ values.

For the proof, see [102].

Remark 5.2.1. Assumption (5.10) is needed for the use of Theorem 5.2.2, in which the condition for the eigenvalues of submatrix $L_{3}$ is only sufficient, but we do not know whether it is necessary (see Remark 10 in [25]).

Now we formulate the theorem that states that for a sufficiently large $D_{T}$ the equilibrium will be stable.

Theorem 5.2.4. Let $D_{P}$ and $D_{R}$ be two fixed positive numbers, and suppose that the following conditions hold:

$$
\begin{align*}
B & <\min \{a+c, A+h\},  \tag{5.11}\\
B\left(A c+a h+(a+c)^{2}\right) & <c^{2}(A+a)+B^{2}(a+c)+a^{2}(c+h) .
\end{align*}
$$

Then there is a positive number $\tilde{D}_{T}$ such that the system (5.4) is stable for all $D_{T}>\tilde{D}_{T}$ values.

For the proof, see [102].

### 5.3 Numerical solution

In this section we solve equation (5.4) numerically, and for this we use the finite element method. First we state the weak form of the problem: we multiply both sides of our equation by a function $v \in H^{1}(\Omega)$, and then integrate it on $\Omega$. In this way we get the equation

$$
\begin{equation*}
\int_{\Omega} v(x) \partial_{t} Q(t, x) \mathrm{d} x=\int_{\Omega} F(Q(t, x)) v(x)+v(x) D_{Q} \Delta Q(t, x) \mathrm{d} x . \tag{5.12}
\end{equation*}
$$

From now on we will search for such a function $Q(t, x) \in H^{1}(\Omega)$ for which 5.12 holds for every $v \in H^{1}(\Omega)$. For the sake of simplicity, we use the usual $L^{2}$ scalar product:

$$
(u, v):=\int_{\Omega} u v .
$$

In this way the previous equation takes the form

$$
\begin{equation*}
\left(\partial_{t} Q(t, x), v(x)\right)=(F(Q(t, x)), v(x))+\left(D_{Q} \Delta Q(t, x), v(x)\right) . \tag{5.13}
\end{equation*}
$$

Using the Gauss-Ostrogradsky theorem and the Neumann boundary condition, we can rewrite the right-hand side in the following way:

$$
\begin{equation*}
\left(\partial_{t} Q(t, x), v(x)\right)=(F(Q(t, x)), v(x))-\left(D_{Q} \nabla Q(t, x), \nabla v(x)\right) . \tag{5.14}
\end{equation*}
$$

Let us search for our numerical solution in a $V_{h} \subset H^{1}(\Omega)$ subspace (here we will use triangular quadratic Lagrange elements) in the form

$$
Q(t, x) \approx \sum_{j} c_{j}(t) b_{j}
$$

where $\left\{b_{j}\right\}_{j=1, \ldots, n}$ is a basis in $V_{h}$. Also, let us choose our function $v$ to be one of these basis functions. Consequently, (5.14) has the form

$$
\begin{equation*}
\left(\partial_{t} \sum_{j} c_{j}(t) b_{j}, b_{k}\right)=\left(F\left(\sum_{j} c_{j}(t) b_{j}\right), b_{k}\right)-\left(D_{Q} \sum_{j} c_{j}(t) \nabla b_{j}, \nabla b_{k}\right), \tag{5.15}
\end{equation*}
$$

for every $k=1,2, \ldots, n$.
Let us denote our mass matrix by $\mathcal{M}$, which has elements

$$
\mathcal{M}_{k, j}:=\left(b_{j}, b_{k}\right),
$$

and our stiffness matrix by $\mathcal{S}$ with elements

$$
\mathcal{S}_{k, j}:=-\left(D_{Q} \nabla b_{k}, \nabla b_{j}\right) .
$$

We also use the notation

$$
F_{k}(c(t)):=\left(F\left(\sum_{j} c_{j}(t) b_{j}\right), b_{k}\right) .
$$

Therefore, equation (5.15) can be rewritten as

$$
\begin{equation*}
\mathcal{M} c^{\prime}(t)=\mathcal{S} c(t)+F(c(t)) . \tag{5.16}
\end{equation*}
$$

This is a system of ordinary differential equations, which can be solved with any standard method; here we use the BDF method. Moreover, quadratic Lagrange elements are used
as a base on a triangular grid. The numerical simulations were conducted in COMSOL.
In the numerical experiments we choose the initial distribution as follows. People are only present near one of the shores of the island, rats are chosen similarly and trees are distributed homogeneously on the island. This initial condition corresponds to the time of the arrival of the settlers. We will use the following values for the parameters in system (5.4), similarly as in Chapter 4

$$
a=0.03, \quad b=1, \quad c=10, \quad M=12000, \quad f=0.001, \quad h=0.25, \quad \mathcal{A}=162 .
$$

Now we examine whether the analytic results can be verified numerically. As we can see in Figure 5.3, for parameters $\left(D_{P}, D_{R}, D_{T}\right)=(0.1,0.3,0.01)$ the system is unstable as the population of the trees dies out where people first start to harvest. Note that in the case of $T=0$ the derivatives in the first two equations will become infinitely large, resulting in the loss of validity of our model. To avoid such problems we will only consider the case of sufficiently large values of $T$ and say that for values near zero the populations will die out.


Figure 5.3: The unstable case - the number of trees at times $t=1$ (upper left), $t=57$ (upper right) and $t=105$ (bottom) for constants $\left(D_{P}, D_{R}, D_{T}\right)=(0.1,0.3,0.01)$. The trees die out at time $t=106$ at the first human settlement.

However, as we increase the value of $D_{T}$, system (5.4) becomes stable e.g. for values
$\left(D_{P}, D_{R}, D_{T}\right)=(0.1,0.3,1)$, as the solution converges to the coexistence equilibrium (5.5) (see Figure 5.4). Note that in this case the sufficient condition in Theorem 5.2.3 does not hold, but Theorem 5.2.4 does hold, i.e. the system may become unstable for certain $D_{T}$ values but will become stable for sufficiently large ones.


Figure 5.4: The stable case - the number of trees at times $t=6$ (upper left), $t=120$ (upper right) and $t=476$ (right) for constants $\left(D_{P}, D_{R}, D_{T}\right)=(0.1,0.3,1)$. Although the number of the trees decreases as time goes by, they do not die out, but rather have a constant positive density on the island, which corresponds to the value which can be computed from (5.5).

### 5.4 Conclusions

In this chapter a two-dimensional extension of the previously discussed equation modeling the ecological collapse of Easter island was introduced. In Section 5.1 the previous onedimensional model was modified into a two (space)-dimensional system of partial differential equations. Then, in Section 5.2 it was shown that the increase of the diffusion of the trees stabilizes the system similarly as in the one dimensional case. Finally, in Section 5.3 the system of partial differential equations was solved numerically using a finite element
method. The numerical results confirmed the previously proved theoretical results.

## Chapter 6

## Possible further extensions

In Chapters 4 and 5, we described a one- and a two-dimensional model describing the ecological collapse of Easter island, respectively. In this chapter, some further possible extensions and observations are mentioned.

In Section 4.1.2 we showed a bound for parameter $f$ above which the system becomes unstable. Similar observations can also be done in the case of the two-dimensional model.

We also know that there is a constant eastern wind on the island, which could be included in our model, even in the one dimensional spatial case. For this, the addition of a convection term would be necessary, and then we can get a partial differential equation. As it was mentioned before, it would also be useful in the two-dimensional case.

It is also important to note that we have only observed the stability of the equilibrium points or the stationary solutions of the systems - however, the dynamics can be understood better if we also search for periodic orbits in the system. Figure 4.13 hints at the existence of such an orbit, so its existence (or stability) might be an interesting question for further research. It is also not rare for three dimensional systems that some other (sometimes even chaotic) orbits might appear, so the observation of such would also be interesting.

Note that this second part of the dissertation focuses on the qualitative properties of the analytic solution, while in the first one numerical methods were examined. It would also be interesting to construct such methods which preserve the properties of the original continuous model (e.g. positivity) and give sufficient bounds for the time steps under which these conditions hold.

## Bibliography

[1] Abbey, H., An examination of the Reed-Frost theory of epidemics, Hum. Biology 24.3 (1952): 201-233.
[2] Akima, H., A new method of interpolation and smooth curve fitting based on local procedures, J. ACM, 17.4 (1970): 589-602. https://doi.org/10.1145/321607. 321609.
[3] Akima, H., A method of bivariate interpolation and smooth surface fitting based on local procedures, Commun. ACM 17 (1974): 18-20. https://doi.org/10.1145/ 360767.360779 .
[4] Aldous, D.J., Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today, Stat. Sci. 16.1 (2001): 23-34. https://doi.org/10.1214/ss/ 998929474 .
[5] Alsenafi, A., Barbaro, A, A convection-diffusion model for gang territoriality, Physica A: Statistical Mechanics and its Applications 510 (2018): 765-786. https: //doi.org/10.1016/j.physa.2018.07.004.
[6] Andrews, L. C., Special functions of mathematics for engineers, Vol. 49. SPIE Optical Engineering Press, Bellingham, WA; Oxford University Press, Oxford, second ed., 1998.
[7] Angrist, M., Here is a Human Being: At the Dawn of Personal Genomics., Harper, 2010.
[8] Aronson, D. G., The asymptotic speed of propagation of a simple epidemic, Nonlinear diffusion 14 (1977): 1-23.
[9] BACAËR, N., A short history of mathematical population dynamics. Springer Science \& Business Media, 2011.
[10] Bagrinovskir, K.A., Godunov, S.K, Difference schemes for multidimensional problems, Dokl. Akad. Nauk. USSR 115 (1957): 431-433.
[11] Bahn, P.G., Flenley, J., Easter Island, Earth Island, Thames and Hudson, New York, 1992.
[12] Bartlett, M. S., Some evolutionary stochastic processes, J. Roy. Statist. Soc. B 11.2 (1949): 211-229.
[13] Basener, W., Brooks, B., Radin, M., Wiandt, T., Dynamics of a Discrete Population Model for Extinction and Sustainability in Ancient Civilizations, Nonlinear Dynamics, Psychology, and Life Sciences 12.1 (2008): 29-53.
[14] Basener, W., Brooks, B., Radin, M., Wiandt, T., Rat Instigated Human Population Collapse on Easter Island, Nonlinear Dynamics, Psychology, and Life Sciences 12.3 (2008): 227-240.
[15] Basener, W., Brooks, B., Radin, M., Wiandt, T., Spatial Effects and Turing Instabilities in the Invasive Species Model, Nonlinear Dynamics, Psychology, and Life Sciences 15.4 (2011): 455-464.
[16] Bellman, R., On the computational solution of differential-difference equations, J. Math. Anal. Appl. 2.1 (1961): 108-110.
[17] Bernoulli, D., Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir Hist. Acad. R. Sci. Paris (1760/1766): 1-45. English translation: Rev. Med. Virol. 14 (2004): 275-288.
[18] Bernoulli, D., Observationes de seriebus... Comment. Acad. Sci. Imp. Petropolitanae 3 (1728/1732): 85-100. Reproduced in: Die Werke von Daniel Bernoulli, Band 2, Birkhäuser, Basel (1982): 49-64.
[19] Berthollet, C-L., Essai de statique chimique, Vol. 1. F. Didot, 1803.
[20] Bonnasse-Gahot, L., Berestycki, H., Depuiset, M-A., Gordon, M. B., Nadal, J-P., Roché, S., Rodríguez., N., Epidemiological Modeling of the 2005 French Riots: A Spreading Wave and the Role of Contagion, Sci. Rep. 8.1 (2018): 1-20.
[21] Butcher, J. C., Numerical methods for ordinary differential equations, John Wiley \& Sons, Ltd., Chichester, third ed., 2016. https://doi.org/10.1002/ 9781119121534
[22] Capasso, V., Fortunato, D., Stability results for semilinear evolution equations and their application $t$ some reaction-diffusion problems, SIAM J. Appl. Math. 39.1. (1980): 37-47. https://doi.org/10.1137/0139005
[23] Carlson, R. E., Fritsch, F. N., Monotone piecewise bicubic interpolation, SIAM J. Numer. Anal. 22.2 (1985): 386-400. https://doi.org/10.1137/0722023.
[24] Carlson, R. E., Fritsch, F. N., An algorithm for monotone piecewise bicubic interpolation, SIAM J. Numer. Anal., 26.1 (1989): 230-238. https://doi.org/10. 1137/0726013.
[25] Casten, R. G., Holland, C. J., Stability properties of solutions to systems of reaction-diffusion equations, SIAM J. Appl. Math, Vol. 33.2. (1977): 353-364. https://doi.org/10.1137/0133023
[26] Collins, F. S., The Language of Life: DNA and the Revolution in Personalized Medicine, Profile Books, 2010
[27] Connell, J.H., On the role of natural enemies in preventing competitive exclusion in some marine animals and in rain forest trees. In: Dynamics of Population. 298. Ed. Den Boer, P.J. and Gradwell, G.R. (1971): 312.
[28] Cooke, K.L., Stability analysis for a vector disease model, Rocky Mountain J. Math. 9.1 (1979): 31--42. https://www.jstor.org/stable/44238836
[29] Cools, R., Kim, K.J., A survey of known and new cubature formulas for the unit disc, Korean J. Comput. and Appl. Math. 7.3 (2000): 477-485.
[30] Courant, R. Hilbert, D., Methoden der mathematischen Physik I, Springer, 1931 (English transl.: Methods of Mathematical Physics, vol. I., Interscience, 1953)
[31] Csomós, P., Faragó, I., Havasi, Á., Weighted sequential splittings and their analysis, Computers Math. Appl. 50.7 (2005): 1017-1031. https://doi.org/10. 1016/j.camwa.2005.08.004
[32] Csomós, P., Faragó, I., Error analysis of the numerical solution of split differential equations, Mathl. Comput. Modelling 48.7-8 (2008): 1090-1106. https: //doi.org/10.1016/j.mcm.2007.12.014
[33] Csomós, P., Takács, B., Operator splitting for space-dependent epidemic model, Appl. Num. Math. 159 (2021): 259-280. https://doi.org/10.1016/j.apnum. 2020.09.010
[34] Daley, D. J., Gani, J., A deterministic general epidemic model in a stratified population, In: Probability, Statistics and Optimisation: A Tribute to Peter Whittle (Kelly, F. P., Ed.), John Wiley, Chichester, 1994. 117-132.
[35] Daley, D. J., Gani, J., Models for the spread of infection via pairing at parties, In: Advances in Applied Probability and Stochastic Processes: Festschrift for Julian Keilson (Shantikumar, G., Sumita, U., Eds), Kluwer: Boston. 1999. 95-113. https://doi.org/10.1007/978-1-4615-5191-1_8
[36] Daley, D. J., Gani, J., Epidemic modelling: an introduction No. 15. Cambridge University Press, 2001.
[37] Davis, P., Polonsky, I., Numerical Interpolation, Differentiation and Integration, in: M. Abramowitz, I. Stegun, Handbook of Mathematical Functions, NBS, (1964): 876-925.
[38] Davis, P. J., Rabinowitz, P., Methods of numerical integration, Dover Publications, Inc., Mineola, NY, 2007. Corrected reprint of the second (1984) edition.
[39] Davis, K., The \$ 1,000 Genome: The Revolution in DNA Sequencing and the New Era of Personalized Medicine, Free Press, 2010
[40] Descombes, S., Duarte, M., Massot, M., Operator splitting methods with error estimator and adaptive time-stepping. Application to the simulation of combustion phenomena. In: R. Glowinski, S. Osher, and W. Yin (eds.): Splitting Methods in Communication, Imaging, Science, and Engineering, Springer International Publishing (2015): 1-13. https://doi.org/10.1007/978-3-319-41589-5_19
[41] Dougherty, R. L., Edelman, A. S., Hyman, J. M., Nonnegativity-, monotonicity-, or convexity-preserving cubic and quintic Hermite interpolation, Math. Comp. 52.186 (1989): 471-494. https://doi.org/10.2307/2008477.
[42] Elhay, S., Kautsky, J., Algorithm 655: Iqpack: Fortran subroutines for the weights of interpolatory quadratures, ACM Trans. Math. Software 13.4 (1987): 399415. https://doi.org/10.1145/35078.214351
[43] Elsgolts, L.E., Qualitative Methods in Mathematical Analysis, Vol. 12. Amer. Math. Soc., Providence R.I., 1964.
[44] Euler, L., Recherches générales sur la mortalité et la multiplication du genre humain Hist. Acad. R. Sci. B.-Lett. Berl. 16 (1760/1767) 144-164.
[45] Faragó, I., Horváth, R., On some qualitatively adequate discrete space-time models of epidemic propagation, J. Comp. Appl. Math. 293 (2016): 45-54. https: //doi.org/10.1016/j.cam.2015.03.030
[46] Faragó, I., Horváth, R., Qualitative properties of the finite difference solution of a space-time epidemic propagation model, Annales Univ. Sci. Budapest., Sect. Comp. 45. (2016): 157-168.
[47] Faragó, I., Horváth, R., Qualitative properties of some discrete models of disease propagation, J. Comput. Appl. Math. 340 (2018): 486-500. https://doi.org/10. 1016/j.cam.2017.09.024.
[48] Friedman, A., Partial differential equations, Holt, Rinehart and Winston, Inc., New York- Montreal, Que.-London, 1969.
[49] Fritsch, F. N., Carlson, R. E., Monotone piecewise cubic interpolation, SIAM J. Numer. Anal. 17.2 (1980): 238-246. https://doi.org/10.1137/0717021.
[50] Fritsch, F. N., Carlson, R. E., Monotonicity preserving bicubic interpolation: a progress report, Computer Aided Geometric Design 2.1-3 (1985): 117-121. https : //doi.org/10.1016/0167-8396(85)90014-7.
[51] Gani, J., Jerwood, D., Markov chain methods in chain binomial epidemic models, Biometrics 27 (1971): 591-603.
[52] Gottlieb, S., Shu, C-W., Total variation diminishing Runge-Kutta schemes, Mathematics of Computation 67.221 (1998): 73-85. https://doi.org/10.1090/ S0025-5718-98-00913-2.
[53] Gottlieb, S., Shu, C-W., Tadmor, E., Strong stability-preserving high-order time discretization methods, SIAM Rev. 43.1 (2001): 89-112. https://doi.org/ 10.1137/S003614450036757X.
[54] Gottlieb, S., Ketcheson, D. I., Shu, C-W., High order strong stability preserving time discretizations, J. Sci. Comput. 38.3 (2009): 251-289. https://doi.org/ 10.1007/s10915-008-9239-z.
[55] Gottlieb, S., Ketcheson, D. I., Shu, C-W., Strong stability preserving RungeKutta and multistep time discretizations, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2011, https://doi.org/10.1142/7498.
[56] Graunt, J. Natural and Political Observations Made Upon the Bills of Mortality John Martin, London, 1662. [Reprinted in facsimile in The Earliest Classics: John Graunt and Gregory King Gregg International Publishers, Farnborough, 1973]
[57] Greene, W. H., Econometric Analysis, Prentice Hall, 2002.
[58] Halley, E., An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw; with an attempt to ascertain the price of annuities upon lives, Phil. Trans. Roy. Soc. London 17 (1693): 596-610. https://doi.org/10.1017/S2046167400044712.
[59] Harko T., Lobo F.S.N., and Mak M.K., Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates, Appl. Math. Comp. 236 (2014): 184-194. https: //doi.org/10.1016/j.amc.2014.03.030.
[60] Hunt, T., Rethinking the fall of Eater Island: New ecidence points to an alternative explanation for a civilization's collapse, Amer. Sci. 94.5 (2006): 412-419.
[61] Hunt, T., Rethinking Eater Island's ecological catastrophe, J. Arch. Sci. 34.3 (2007): 485-502. https://doi.org/10.1016/j.jas.2006.10.003.
[62] Hunt, T., Lipo, C., The Statues Walked - What Really Happened on Easter Island, http://longnow.org/seminars/02013/jan/17/ statues-walked-what-really-happened-easter-island/ Last accessed 14 October 2021
[63] Hurwitz, A., Ueber die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt, Math. Ann. 46.2 (1895): 273-284. https: //doi.org/10.1007/BF01446812.
[64] Janzen, D. H., Herbivores and the Number of Tree Species in Tropical Forests. Amer. Nat. 104.940 (1970): 501-528. https://doi.org/10.1086/282687
[65] Kautsky, J., Elhay, S., Calculation of the weights of interpolatory quadratures, Numer. Math. 40.3 (1982): 407-422. https://doi.org/10.1007/BF01396453.
[66] Keliger, D., Horváth, I., Takács, B., Local-density dependent Markov processes on graphons with epidemiological applications, Stoch. Proc. Appl., accepted. https: //arxiv.org/abs/2008.08109
[67] Kendall, D.G., in discussion on Bartlett, M.S., Measles periodicity and community size, J. Roy. Stat. Soc. Ser. A 120.1 (1957): 48 - 70. https://doi.org/10.2307/ 2342553
[68] Kendall, D.G., Mathematical models of the spread of infection, In : Mathematics and Computer Science in Biology and Medicine. H.M.S.O. (1965): 213 - 225.
[69] Kermack, W.O., McKendrick, A.G., A contribution to the mathematical theory of epidemics, Proc. R. Soc. A: Math. Phys. Eng. Sci. 115.772 (1927): 235-240. https://doi.org/10.1098/rspa.1927.0118.
[70] Laplace, P.S., Théorie analytique des probabilités, Courcier, Paris, 1812.
[71] Lotka, A.J., Relation between birth rates and death rates, Science 26.653 (1907): 21-22. Reprinted in Smith \& Keyfitz (1977)
[72] Lotka, A.J., Analytical note on certain rhythmic relations in organic systems, Proc. Natl. Acad. Sci. 6.7 (1920): 410--415. https://doi.org/10.1073/pnas.6.7.410.
[73] Lotka, A.J., Undamped oscillations derived from the law of mass action, J. Amer. Chem. Soc. 42.8 (1920): 1595-1599. https://doi.org/10.1021/ja01453a010.
[74] LovÁsZ, L., Large networks and graph limits, Vol. 60. Amer. Math. Soc., 2012.
[75] Ma, J., Rokhlin, V., Wandzura, S., Generalized Gaussian quadrature rules for systems of arbitrary functions, SIAM Numer. Anal. 33.3 (1996): 971-996. https: //doi.org/10.1137/0733048.
[76] Malthus, T.R., An Essay on the Principle of Population, 1st edn. London, 1798.
[77] Marchuk, G.I., Some application of splitting-up methods to the solution of mathematical physics problems, Applik. Mat. 13.2 (1968): 103-132. https://doi.org/ 10.21136/AM.1968.103142
[78] Martin, R. S., Wilkinson, J. H., Handbook Series Linear Algebra: The implicit QL algorithm, Numer. Math. 12.5 (1968): 377-383. https://doi.org/10.1007/ BF02161360.
[79] McKendrick, A.G., Applications of mathematics to medical problems, Proc. Edinb. Math. Soc. 13. (1926): 98-130.
[80] Miller, J.C., A note on the derivation of epidemic final sizes, Bull. Math. Biol. 74.9 (2012): 2125-2141. https://doi.org/10.1007/s11538-012-9749-6.
[81] Miller, J.C., Mathematical models of SIR disease spread with combined non-sexual and sexual transmission routes, Inf. Dis. Model. 2.1 (2017) 35-55. https://doi.org/ 10.1016/j.idm.2016.12.003
[82] Morris, J., The Routh and Routh-Hurwitz stability criteria, Aircraft Eng. 34.395 (1962): 25-27. https://doi.org/10.1108/eb033509.
[83] Othmer, H. G., Scriven, L. E., Interactions of reaction and diffusion in open systems, Ind. Eng. Chem. Fundamen. 8.2 (1969): 302--313. https://doi.org/10. 1021/i160030a020.
[84] Perumpanani, A. J., Sherratt, J. A., Maini P. K., Phase differences in reaction-diffusion-advection systems and applications to morphogenesis, IMA J. Appl. Math. 55.1 (1995): 19-33. https://doi.org/10.1093/imamat/55.1.19.
[85] Reed, M. C., Mathematical biology is good for mathematics., Notices of the AMS 62.10 (2015): 1172-6.
[86] Rendine, S., Piazza, A., Cavalli-Sforza, L. L., Simulation and separation by principal components of multiple demic expansions in Europe, Amer. Nat. 128.5 (1986): 681-706. https://doi.org/10.1086/284597.
[87] Ross, R., The Prevention of Malaria, 2nd edn. John Murray, London, 1911.
[88] Routh, E. J., A Treatise on the Stability of a Given State of Motion: Particularly Steady Motion, Macmillan, 1877.
[89] Routh, E. J., Advanced Rigid Dynamics, Macmillan \& Co., 1892.
[90] Sebestyén, G., Faragó, I., Invasive species model with linear rat harvesting on Easter Island, J. Appl. Computat. Math. 4 (2015): 1-6.
[91] Sharpe, F.R., Lotka, A.J., A problem in age-distribution, Philos. Mag. 6.21 (1911): 435--438. Reprinted in Smith \& Keyfitz (1977) https://doi.org/10. 1007/978-3-642-81046-6_13.
[92] Shu, C.-W., Total-variation-diminishing time discretizations, SIAM J. Sci. Statist. Comput. 9.6 (1988): 1073-1084. https://doi.org/10.1137/0909073
[93] Shu, C.-W., Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws, in: Advanced numerical approximation of nonlinear hyperbolic equations, vol. 1697 of Lecture Notes in Mathematics, Springer, Berlin, (1998) 325-432. https://doi.org/10.1007/BFb0096355.
[94] Shu, C.-W., Osher, S., Efficient implementation of essentially nonoscillatory shock-capturing schemes, J. Comput. Phys. 77.2 (1988): 439-471. https://doi. org/10.1016/0021-9991(88)90177-5.
[95] Sigler, L., Fibonacci's Liber Abaci: a translation into modern English of Leonardo Pisano's book of calculation, Springer Science \& Business Media, 2003.
[96] Strang, G., On the construction and comparison of difference schemes, SIAM J. Num. Anal. 5.3 (1968): 506-517. https://doi.org/10.1137/0705041.
[97] Stroud, A. H., Approximate calculation of multiple integrals, Prentice-Hall, 1971.
[98] Takács, B., Analysis of some characteristic parameters in an invasive species model, Annales. Univ. Sci. Buapest., Sect. Comp. 45 (2016) 119-133. http://ac. inf.elte.hu/Vol_045_2016/119_45.pdf
[99] Takács, B., Horváth, R., Faragó, I., The effect of tree-diffusion in a mathematical model of Easter Island's population, Electron. J. Qual. Theory Differ. Equ. 2016.84 (2016): 1-11. http://real.mtak.hu/45097/1/p5296.pdf.
[100] TakÁcs, B., Modeling the ecological collapse of Easter Island, Master's thesis, 2017.
[101] TakÁcs, B., A Continuous Model for the Ecological Collapse of Easter Island, In: Dimov, I., Faragó, I., Vulkov, L. (eds) Finite Difference Methods. Theory and Applications. FDM 2018. Lecture Notes in Computer Science, vol 11386. Springer, Cham. 525-532. https://doi.org/10.1007/978-3-030-11539-5_61.
[102] Takács, B., Horváth, R., Faragó, I., The effect of tree diffusion in a twodimensional continuous model for Easter Island, Eur. J. Math. 5.3 (2019): 845-857. https://doi.org/10.1007/s40879-018-0276-3.
[103] TakÁcs, B., Horváth, R., Faragó, I., Space dependent models for studying the spread of some diseases, Comp. Math. Appl. 80.2 (2020): 395-404. https://doi. org/10.1016/j.camwa.2019.07.001.
[104] Takács, B., Hadjimichael, Y., High Order discretization methods for spatialdependent SIR models, Math. Comp. Sim., submitted. https://arxiv.org/abs/ 1909.01330
[105] Takács, B., Faragó, I., Horváth, R., Repovš, D., Qualitative properties of space-dependent SIR models with constant delay and their numerical solutions, Computational Methods in Applied Mathematics, submitted. https://arxiv.org/ abs/2112.06808
[106] Thieme, H.R., A model for the spatial spread of an epidemic, J. Math. Biol. 4.4 (1977): 337--351. https://doi.org/10.1007/BF00275082.
[107] Trefethen, L.N., Is Gauss quadrature better than Clenshaw-Curtis?, SIAM Rev. 50.1 (2008): pp. 67-87. https://doi.org/10.1137/060659831.
[108] Verhulst, P.-F., Notice sur la loi que la population poursuit dans son accroissement, Corresp. Math. Phys. 10 (1838): 113--121.
[109] Volkening, A., Linder, D. F., Porter, M. A., Rempala, G. A., Forecasting elections using compartmental models of infections, SIAM Review 62.4 (2020): 837865. https://doi.org/10.1137/19M1306658.
[110] Volterra, V. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi, Mem. Accad. Lincei 6 (1926): 31-113. Reprinted in: Opere matematiche, vol. 5, Accademia nazionale dei Lincei, Roma, 1962.
[111] Volterra, V., Fluctuations in the abundance of a species considered mathematically, Nature 118.2972 (1926): 558--560. Reprinted in L.A. Real, Brown, J.H. (eds.) Foundations of Ecology, University of Chicago Press (1991): 283-285. https://doi.org/10.1038/119012b0.
[112] Waage, P., Guldberg, C. M., Studier over affiniteten, Forhandlinger i Videnskabs-selskabet i Christiania 1 (1864): 35-45.
[113] WaAge, P., Forsøg til Bestemmelse af Lovene for Affiniteten, Forhandlinger i Videnskabs-selskabet i Christiania (1864): 92-94.
[114] Xu, R., Ma, Z., Global stability of a SIR epidemic model with nonlinear incidence rate and time delay, Nonlin. Anal. RWA 10.5 (2009): 3175-3189. https://doi.org/ 10.1016/j.nonrwa.2008.10.013.
[115] Yule, G.U., A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis F.R.S., Phil. Trans. Roy. Soc. Lond. B 213 (1925): 21-87. https: //doi.org/10.1098/rstb. 1925.0002

