Technical University Munich

Faculty of Mathematics

Bachelorthesis
Bsc Mathematik

Pseudomonas Syringae Growth in the Phyllosphere: Revisiting mathematical modelling approaches

Lion Corvin Wallner

Prüferin:	Prof. Dr. Christina Kuttler
Betreuerin:	Dr. Judith Pérez-Velázquez
Abgabetermin:	14.3.2014

Preface

Zusammenfassung

Vorliegende Bachelorarbeit basiert hauptsächlich auf dem stochastischem Modell, das in dem Artikel [1] vorgestellt wird. Es werden zwei verschiedene mathematische Modelle, die sich mit dem Wachstum des *Pseodomonas syringae*, einem Krankheitserreger von Pflanzen, beschäftigen, vorgestellt und überarbeitet.
Die Arbeit ist unterteilt in vier Teile, der erste enthält eine Erörterung biologischer Hintergründe und eine Einführung in die stochastische Modellierung. Der zweite Teil stellt besagtes Modell, die Implementierung des Autors und eine alternative, räumliche Modellierung vor. Anschließend wird das stochastische Modell anaylisiert und Verbesserungansätze gemacht. Im letzten Teil befinden sich eine Zusammenfassung der Ergebnisse und ein Zukunftsausblick.

Abstract

This bachelor thesis is based on a stochastic model presented in the article [1]. We revise two different mathematical model of growth for the bacteria *Pseudomonas syringae*, a plant pathogen.
The work is divided in four parts, the first explains the biological setting and basics of stochastic modeling. Then, the model, the author's implementation and an alternative spatial model are presented. In the third part the analysis of the model and improvement suggestions can be found. The final part contains a summary and a perspective on future work.

Erklärung zur Bachelorarbeit

Hiermit versichere ich, dass die vorliegende Arbeit von mir selbstständig verfasst wurde und dass keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden. Diese Erklärung erstreckt sich auch auf in der Arbeit enthaltene Graphiken, Zeichnungen, Kartenskizzen und bildliche Darstellungen.

.....	.....
Datum und Ort	Name

Contents

I. Introduction 1

1. Pseudomonas Syringae 1
2. Stochastic Modelling of Bacterial Growth 1
3. Gillespie's Stochastic Simulation Algorithm 4

II. Models and Impementations 7

4. A Stochastic Model of Bacterial Growth 7
5. Numerical Implementation 8
6. PHYLLOSIM - a spatial approach 10

III. Analysis and Improvement of the stochastic Model 12

7. Migration Rate 12
8. Capacity 15
9. Inoculation 17

IV. Conclusion 20

10. Summary and Overview 20
11. Future Work 21

List of Figures

1. Schematic of stochastic simulation algorithm 4
2. Simulations of linear birth-death model 5
3. Exert from the PHYLLOSIM-article 10
4. while-loop of `pcod1.m` 12
5. Examplary Migration probability h(N) 13
6. Simulations with different switchtimes compared 14
7. Examplary plots of migration probability h(N) for different migration rates . . . 15
8. Probability behaviour for different a-values 16
9. Qualitative behaviour of birth rate and event probablities 17
10. Simulation (20,000 events) of required time to reach a population size of 1000 . . 18
11. Behaviour of birth probabilities for differing birth rates 18
12. Examplary initial size distributions 19
13. Visualisation of $beta(\nu, \eta)$ and $unif(0, 1)$ 20

List of Tables

1. Model parameters 8
2. Experimental data by Dulla & Lindow 10
3. Development of $X_{brd} = 2^a$ 16
4. Birth probabilites for bN and $\widetilde{bN}$ 17
5. Input parameters and meaning 21
6. Outline of MATlab-Functions 22

Part I.
Introduction

1. Pseudomonas Syringae

Due to factors like fluctuating temperatures or humidity the phyllosphere (*the surface of the leaf*) considered to be a rather complex environment. However, leaf surfaces still provide a habitat for several microbes, among them *Pseudomonas syringae*.
This is a gram negative bacterium capable of growing on the phyllosphere and damaging leaves (e.g.*Pseudomonas syringae* can cause frost injuries by catalysing ice nucleation), it can also spread plant diseases, potentially destroying vast amounts of crops. As it can live and populate symptomless, environmental changes may cause quick outbreaks with devastating effects. To further understand and investigate its spreading behaviour one can work with mathematical (in particular stochastic) models.

source: [1]

2. Stochastic Modelling of Bacterial Growth

Assuming that bacterial movements or birth-death processes are completely predictable, as done in deterministic models, in general seems not suitable. A stochastical model which assigns probabilities to certain events, e.g. the following by *J.Perez-Velazquez et.al.*, is more realistic and provides several advantages.
Deterministic models, like a continous linear birth-death process , described by the following ODE,

$$\begin{gathered} dN(t)/dt = (\lambda - \mu)N(t) \\ (N(t) : population\ size,\ \lambda :\ birth\ rate,\ \mu :\ death\ rate) \end{gathered} \tag{1}$$

can take non-integer values for N(t), which should be a natural number. Also, deterministic models do not produce fluctuations, e.g. in the linear case: if the net rate $(\lambda - \mu)$ equals zero, the population size $N(t)$ remains constant, unlike real data.
These fluctuations in stochastic models can lead to totally different results, although initial states remain unchanged for several realizations and stochastic models cannot be reproduced.
The corresponding stochastic model to the deterministic linear birth-death process is

$$dP_N(t)/dt = \lambda(N-1)P_{N-1}(t) - (\lambda+\mu)NP_N(t) + \mu(N+1)P_{N+1}(t) \tag{2}$$

($P_N(t)$ denotes the probability that the population size equals N at time t). Its solution can be the coefficient of z^N in the expansion of

$$\left\{ \frac{\mu(1-z) - (\mu - \lambda z)exp(-(\lambda-\mu)t)}{\lambda(1-z) - (\mu-\lambda z)exp(-(\lambda-\mu)t)} \right\}^{n_0}$$

(n_0 is the initial population size), but handling this expression becomes impractical for $n_0 > 1$. Further analysis could be made but this thesis will focus on simulation.
The following is a rough introduction to chemical reactions systems and short explanation of the mathematical background of the stochastic simulation algorithm by Daniel T. Gillespie, Fig.**??** , for more information consult [5].
Most general chemical reactions with n chemical species $X_{1,\dots,n}$ and m chemical reactions $R_{1,\dots,m}$ are of the form:

$$\nu_{1,j}X_1 + \nu_{2,j}X_2 + ... + \nu_{n,j}X_n \underset{k_{-j}}{\overset{k_{+j}}{\rightleftharpoons}} \eta_{1,j}X_1 + \eta_{2,j} + ... + \eta_{n,j}X_n \tag{3}$$

with $i \in [n],\ j \in [m]$:

- $\nu_{i,j}, \eta_{i,j} \in \mathbb{N}_0$ are stoichiometric coefficients
- $k = [k_{+1}, k_{-1}, k_{+2}, k_{-2}, ..., k_{+m}, k_{-m}] \in \mathbb{R}_+^{2m}$ are the deterministic reaction rate constants.

For k_{+j}-rates the lefthand side of the chemical reaction R_j contains the reactant molecules.[1]
Let $x = [x_1,\ x_2,\ ...,\ x_n]^T$ be the concentration state vector, $x_i, i \in \mathbb{N}$ is the concentration of the chemical species X_i. $x_i = \frac{X_i}{\Omega}$, where Ω denotes the reaction volume. X_i is the number of molecules in Ω.
The reaction flux $v_j(x) = k_{+j} \prod_{i=1}^n x_i^{\nu_{i,j}} - k_{-j} \prod_{i=1}^n x_i^{\eta_{i,j}}$ is the instantaneous frequency with which reaction R_j takes place given a concentration state x. v_j consists of $k_{+j} \prod_{i=1}^n x_i^{\nu_{i,j}}$, the forward reaction, and $k_{-j} \prod_{i=1}^n x_i^{\eta_{i,j}}$, the backward reaction.
The stoichiometric coefficients $\nu_{i,j}, \eta_{i,j}$ lead to the stoichiometric matrix S:

$$S = \begin{pmatrix} \eta_{1,1} - \nu_{1,1} & \eta_{1,2} - \nu_{1,2} & \dots \\ \eta_{2,1} - \nu_{2,1} & \eta_{2,2} - \nu_{2,2} & \dots \\ & \vdots\,\vdots & \end{pmatrix}.$$

The number of rows equals n, the number of columns equals m. We can define an ODE model of this system, also called the reaction rate equation, $v = [v_1,\ v_2,\ ...,\ v_m]^T$:

$$\dot{x} = Sv(x).$$

Note that this is still a deterministic model. (x_i are concentrations, k contains rates) A short demonstration with the deterministic linear birth-death model, which is equal to the following:

$$\begin{array}{lll} X & \underset{0}{\overset{\lambda}{\rightleftharpoons}} & 2X \\ X & \underset{0}{\overset{\mu}{\rightleftharpoons}} & \emptyset \end{array} . \qquad (4) \qquad \begin{array}{ll} \lambda\text{:} & \text{birth rate} \\ \mu\text{:} & \text{death rate} \end{array}$$

As only forward reactions exist ($k_{-1} = k_{-2} = 0$) the flux vector $v = \begin{pmatrix} \lambda x \\ \mu x \end{pmatrix}$ is very simple (the right hand side of the reaction equation is meaningless for the evaluation of v). With $S = (2 - 1 \quad 0 - 1)$ this leads to

$$\dot{x} = Sv(x) = (1 \quad -1) \begin{pmatrix} \lambda x \\ \mu x \end{pmatrix} = (\lambda - \mu)x. \qquad (5)$$

x is equal to a concentration ($x = \frac{X}{\Omega}$). Ω can be treated as a constant. Multiplying Equation (5) with Ω procudes an equation similiar to the deterministic linear birth-death process:

$$\dot{X} = (\lambda - \mu)X.$$

The stochastic reaction constant $c_\mu dt$, defined as the average probability that a particular combination of R_μ reactant molecules will react accordingly in the next infinitesimal time interval dt, is closely related to the deterministic reaction-rate constant k_μ. D. T. Gillespie's approach reaches the form

[1] Chemical reactions R_μ with $k_{+\mu}, k_{-\mu} > 0$ can be regarded as two reactions with no backward reaction. Then it is possible to arrange them, so only lefthand-sides contain the according reactants which can be easier to handle. In the following, if not stated otherwise, we let reactions be in that form.

$$k_\mu = \Omega^{N-1} \frac{\langle \prod_{i=1}^{N} X_i \rangle}{\prod_{i=1}^{N} \langle X_i \rangle} c_\mu.$$

for chemical reactions which require at most one of each reactant (that means $\nu \in 0, 1$), where N is the number of reactants.[2] Gillespie does not distinguish between the product of averages and the average of a product (*Detemernistic formulation of chemical kinetics*) and acquires the form:

$$k_\mu = \Omega^{N-1} c_\mu.$$

In the deterministic linear birth-death model the deterministic and stochastic reaction rate are equal ($k_\mu = c_\mu$), because it has only one reactant molecule (see (4)).
Gillespie uses the *reaction probability density function*[3] $P(\tau, \mu)$ to develop his algorithm:

$$\begin{aligned} P(\tau,\mu)d\tau \quad \equiv \quad & \textit{probability that given the state } (X_1, ..., X_N) \textit{ at time } t \textit{ the} \\ & \textit{next reaction } V \textit{ will occur in the } infinitesimal \textit{ time} \\ & \textit{interval } (t+\tau, t+\tau+d\tau) \textit{ and will be an } R_\mu \textit{ reaction} \end{aligned}$$

and $P_0(\tau)$ shall be the probability that no reaction will take place in $(t, t+\tau)$ for a given state $(X_1, ..., X_N)$ and time t. Furthermore we define:

$$\begin{aligned} h_\mu \quad \equiv \quad & \textit{number of distinct } R_\mu \textit{ molecular reactant combinations} \\ & \textit{available in the state } (X_1, ..., X_N). \end{aligned}$$

The stochastic reaction constant and h_μ lead to:

$$\begin{aligned} a_\mu dt \quad \equiv \quad h_\mu c_\mu dt \quad \equiv \quad & \textit{probability that an } R_\mu \textit{ reaction will occur in } V \textit{ in } (t, t+dt) \\ & \textit{in the state } (X_1, ..., X_N) \textit{ at time } t \end{aligned}$$

The *reaction probability density function* can now be expressed as:

$$P(\tau, \mu)d\tau = P_0(\tau) a_\mu d\tau. \tag{6}$$

With $P_0(\tau' + d\tau') = P_0(\tau')(1 - \sum_{\nu=1}^{m} a_\nu d\tau')$, the probability that no reaction will occur in time $d\tau'$, given the state $(X_1, ...X_N)$ is $(1 - \sum_{\nu=1}^{m} a_\nu d\tau')$, providing us: $P_0(\tau) = exp(-\sum_{\nu=1}^{m} a_\nu d\tau)$
Thus, (6) evolves to ($a_0 \equiv \sum_{\nu=1}^{m} a_\nu$ is also known as "total reaction rate"):

[2] A reaction which needs more than 1 molecule of one chemical species, e.g. $R_* : \quad 2X \rightarrow anything$, has the relation $k_* = Vc_*/2$. There are $X(X-1)/2!$ possibilities to match molecule X with another X, not XX as a molecule cannot match itself.

[3] Gillespie also uses and explains more thoroughly with aid of the *chemical master equation*. Consult [5] for more information.

$$P(\tau,\mu) = \begin{cases} a_\mu exp(-a_0\tau) & if\ r \in [0,\infty), \mu \in \{1,...,m\} \\ 0 & else \end{cases} \tag{7}$$

Overall, $P(\tau,\mu)$ in Equation (7) requires:

- **all** m reaction constants $c_1,...,c_m$
- the current molecule number of **all** reactant species $X_1,...,X_N$

Depending on the case one can speak of events instead of reactions, e.g. if population events like births and deaths are regarded.

sources:[2],[4],[5]

3. Gillespie's Stochastic Simulation Algorithm

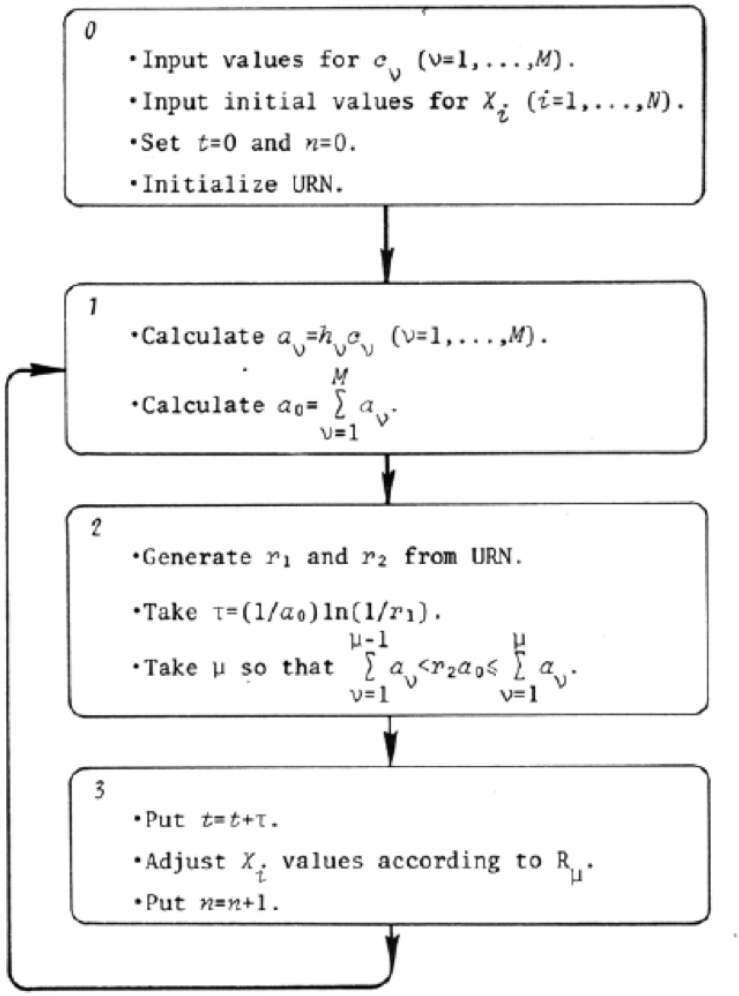

Figure 1: Schematic of stochastic simulation algorithm

N : *number of species* c_ν : *stochastic reaction constant*
M : *amount of reaction equations* a_ν : *reaction rates*

source: see [5]

To implement numerically a stochastic process starting from a given state, the needed informations are the interevent-time, to say the duration until the next event occurs, and an entity

telling what kind of event happens then. Gillespie describes in Fig.1 an algorithm that simulates both requirements with the help of two unit-interval uniform random numbers with three steps (and an initialization). The two random variables shall be denoted as $r_1, r_2 \in (0,1)$ respectively.[4] The general idea is that from these random numbers the interevent-time τ and kind of event μ can be generated corresponding to the *reaction/event probability density function.*

$\tau = \frac{1}{a_0} ln(\frac{1}{r_1})$ and μ is natural so $\sum_{\nu=1}^{\mu-1} a_\nu < r_2 a_0 \leq \sum_{\nu=1}^{\mu} a_\nu$.[5]

Following the schematic in Fig.1, for each loop a *unit-interval uniform random number generator* (URN) creates new variables r_1, r_2. That means that even with the same initial conditions (and unchanged r_1, r_2) in *step 0* multiple realisations will most likely produce different results which can, obviously depending on parameters and the model, disperse greatly. In practice one will run the algorithm until the number of simulated events n or the elapsed simulation time t reach a certain limit. A *total event rate* a_0 equaling zero is an abort criterion, a zero-divison in *step 2* must be avoided.

If we develop the example of the linear birth-death model (expressed as a chemical reaction in (4)) further, then:

$$\begin{array}{llll} c_1 = \lambda & h_1 = X & \Rightarrow & a_1 \equiv \lambda X \\ c_2 = \mu & h_2 = X & \Rightarrow & a_2 \equiv \mu X \end{array} \quad \Rightarrow \quad a_0 = \lambda X + \mu X \ . \tag{8}$$

This provides for a given state $x = X$ and $\tau \geq 0$:

$$P(\tau, 1) = \lambda x e^{(-(\lambda+\mu)x\tau)}, \quad P(\tau, 2) = \mu x e^{(-(\lambda+\mu)x\tau)} \ . \tag{9}$$

$P(\tau, 1)$ can be interpreted as the probability that a birth takes place after an elapsed time τ, $P(\tau, 2)$ for a death respectively.

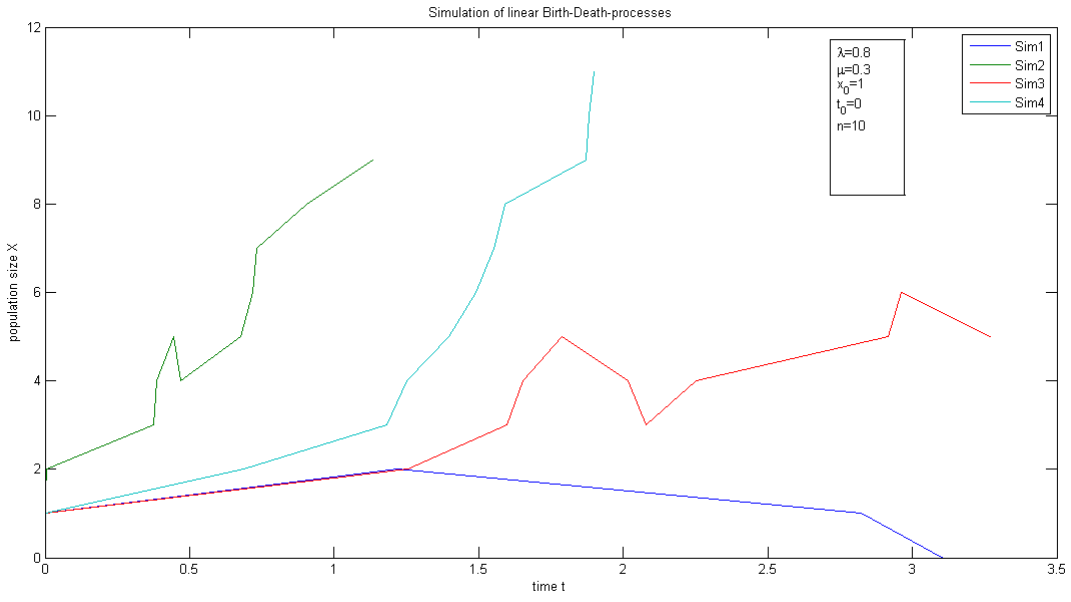

Figure 2: Simulations of linear birth-death model $\lambda = 0.8,\ \mu = 0.3$, 10 events

Fig.2 presents four exemplary run simulations for the linear birth-death model, generated by *SimLinBirthDeath.m*. Each realisation generated ten events starting at time $t_0 = 0$ with birth rate $\lambda = 0.8$ and death rate $\mu = 0.3$ and an initial population $x_0 = 1$.

[4] The boundaries are excluded to prevent problems with zero-division.

[5] With aid of an independence assumption $P_1(\tau)P_2(\mu) = P(\tau, \mu)$, $\tau \in \mathbb{R}$ and $\mu \in \mathbb{N}$ are generated according to $P_1(\tau) = a_0 e^{(-a_0\tau)}$ and $P_2(\mu) = \frac{a_\mu}{a_0}$.

Notice that $Sim1$ and $Sim3$ do not reach values as high as $Sim2$ and $Sim4$ and events happen less frequently (in the case $Sim1$ the third event already leads to extinction, but requires almost the doubled time $Sim4$ needs to complete ten events). This is connected with the *total event rate* a_0, decreasing for lower values of X, thus the *interevent-time* $\tau = \frac{1}{a_0} ln(\frac{1}{r_1})$ in *step 2* tends to larger values for smaller populations.

source:[5]

Part II.
Models and Impementations

4. A Stochastic Model of Bacterial Growth

Pérez-Velázquez et al. consider a continuous-time Markov process $A(t)$ to simulate the bacterial colony dispersal of *Pseudomonas Syringae* on the phyllosphere:

$$A(t) = (N(t), X_1(t), X_2(t), ...), t > 0 \tag{10}$$

where $N(t)$ is the number of colonies that were formed until time t and $X_1(t), X_2(t), ...$ stand for the population sizes of the aggregates. The inititial state is $A(0) = (1, 0, 0, 0, ...)$. *Pérez-Velázquez et al.* state that linear birth-death process do not provide sufficient quality compared to reported distributions, they relate on a logistic birth-death process:
Let $X(t)$ be a markov process with state space $S = \{0, 1, ..., K\}, K \in \mathbb{N}$ displays the carrying capacity.

Definition: $\lambda_i := i\lambda(1 - \frac{i}{K})$, $\mu_i := i\mu$ with $\lambda > 0, \mu \geq 0$. Then $X(t)$ is called a **logistic birth-death process** with birth rate λ_i and death rate μ_i if and only if for $j \in \{0, ..., K\}$ and $k \in \mathbb{N}\backslash\{j-1, j, j+1\}$ is satisfied:

- $P(X(t + \Delta t) = j + 1 | X(t) = j) = \lambda_j \Delta t$
- $P(X(t + \Delta t) = j - 1 | X(t) = j) = \mu_j \Delta t$
- $P(X(t + \Delta t) = k | X(t) = j) = o(\Delta t)\ as\ \Delta t \to 0$

However, this kind of process is still not able to explain new colonies that were observed on the leaf after four days after the inoculation (*placement of bacteria on the phyllosphere*). As it is known that *Pseudomonas Syringae* is capable of movement on the leaf surface *Pérez-Vélazquez et al.* presume cell migration with the following assumptions:

- Cells land, attach and then move
- Cells can leave the aggregate at any time point
- Cells disperse successively, no group movement!
- Emigration(i.e.*migrating cells abandon leaf*)
- Migrating cells create new colonies, they do not join existing ones
- Colonies are created by inoculation or migration
- Inoculation happens only once (no additional immigration)
- Each colony has a random capacity (log-normally distributed)
- Migration doesn't depend on the aggregate size, it happens at a constant rate

The final model of *Pérez-Velázquez et al.* is a **stochastic logistic birth-death process with migration** with the following parameters and transition probabilites:

Table 1: Model parameters

Parameter	Description
$\lambda(=0.4)$	Growth rate
$\mu(=0.1)$	Death rate
I	Migration rate
ρ, σ	lognormal parameters

$$\dot{p}_x^{X_i}(t) = p_{x-1}^{X_i}(t)\lambda_{i,x-1} + (\mu_{x+1} + I)p_{x+1}^{X_i}(t) - (\lambda_{i,x} + \mu_x + I)p_x^{X_i}(t) \tag{11}$$

$$\dot{p}_m^{N}(t) = p_{m-1}^{N}(t)(m-1)I - p_m^{N}(t)mI \tag{12}$$

where $p_m^N(t) = P(N(t) = m)$ denotes the probability that there are m colonies at time t, $p_x^{X_i}(t) = P(X_i(t) = x)$ the probability that the i-th colony has x cells respectively. The birth, death and migration rate values in Table 1 $\lambda = 0.4,\ \mu = 0.1$, and $I = 0.1$ shall be referred to as standard parameter values.
$\lambda_{i,x} = x\lambda(1 - \frac{x}{K_i})$ is the logistic birth rate, $\mu_x = \mu x$ the linear death rate and $K_i \sim ln\mathcal{N}(\rho, \sigma)$ is the capacity of the i-th colony.[6] Note that (11), (12) are not defined for $x = 0$ or $m = 0$, for this case:

$$\dot{p}_0^{X_i}(t) = (\mu + I)p_1^{X_i}(t)$$
$$\dot{p}_0^{N}(t) = -Ip_0^{N}(t)$$

It is assumed that migration does not alter the size of the affected colony thereby enabling independence assumptions significant for analysis. This leads to the model, $N, Y_1, Y_2, Y_3, ...$ where $Y_i,\ i \in \mathbb{N}$ are independent stochastic processes:

N is linear birth process$(N(0) = 1)$, representing the evolution of the number of colonies

$$\dot{p}_n^{N}(t) = p_{n-1}^{N}(t)(n-1)I - p_n^{N}(t)nI$$

Y_i is a birth-death process$(Y_i(0) = 1)$, representing the *time-shifted* size evolution of the colony i after its creation!

$$\dot{p}_x^{Y_i}(t) = p_{x-1}^{X_i}(t)\lambda_{i,x-1} + (\mu_{x+1})p_{x+1}^{Y_i}(t) - (\lambda_{i,x} + \mu_x)p_x^{Y_i}(t)$$

The process $X_i(t) = \begin{cases} 0 & if\ t < T_i \\ Y_i(t - T_i) & if\ t \geq T_i \end{cases}$, with $T_i = min(t \geq 0 : N(t) \geq i)$ as the formation time of colony i. X_i describes the size development of the i-th colony from initial time 0 onwards.

source: [1]

5. Numerical Implementation

This section explains the structure of my implementation. It can reproduce results from the code from *Pérez-Velázquez et al.*, examine the folder 'PartTwo-Numerical Implementation' for more information on this matter.

[6] The migration constant does not depend on the colony size. See chapter Analysis for more details.

My implementation of this model in Matlab is based on the function `LBD_multiple_wanted_2MigRates.m`. As input values it requires in the following order[7]:

- Birth rate λ
- Death rate μ
- First migration rate $I1$
- Second migration rate $I2$
- Capacities for initial colonies $K*$
- Sizes for inital colonies x_0*
- Initial time t_0
- Wanted time points $wantedTimes*$

All arguments are of the type double, arguments marked with * can be an array, x_0, K naturally must have the same dimension, here s.[8] Let l be the length of $wantedTimes$. Its outputs are $t1, x1 \in \mathbb{R}^{l\times s}$ and $t2, x2 \in \mathbb{R}^{l\times m}$, where m is the number of colonies created by migration.
Columns of $x1$ hold the colony size for each starting colony at time points given by *wanted times*, columns of $x2$ for the colonies formed by migration respectively. $t1, t2$ denote the times corresponding $x1, x2$ and are almost dispensable as they obviously resemble the entries of $wantedTimes$.
`LBD_multiple_wanted_2MigRates.m` uses `LBD_single_wanted.m`, which simulates the events for a single colony. Besides using only one migration rate and scalar values for K and x_0 it has the same kind of input values.
`LBD_single_wanted.m` produces column vectors t_{out}, x_{out}, combined they form the output of `LBD_multiple_wanted_2MigRates.m`. An additional output $migVek$ contains the times migration events occured (i.e. the formation times of migration colonies).
Pérez-Velázquez et al. fix the birthrate $\lambda = 0.4$, death rate $\mu = 0.1$, the migration rates $I1 = I2$, the lognormal-parameters $\rho = 1.4511, \sigma = 1.7025$ for the capacities, $t0 = 0$ for the initial time and the initial sizes of all colonies to 1. Their implementation `callpod.m` requires the user to enter the number of realistations, i.e. the number of starting colonies, here denoted by s, and a migration rate I. Running `LBD_multiple_wanted_2MigRates(0.4,0.1,I,I,CapacityGenerator(s),ones(s),0,wantedTimes)` is an equivalent realisation, but only requested data will be given out.
Setting $wantedTimes = [24\ 48\ 72\ 96]$ will produce data $t1, x1, t2, x2$ so `DataOverview.m` provides a possibility to transform the composed vector $x = [x1\ x2]$ enabling comparison with the data by Dulla and Lindow, Table 2. Henceforth $wantedTimes = [24\ 48\ 72\ 96]$ if not stated otherwise.
As this is a stochastic simulation one is well advised to conduct several realisations. Executing `getSimulationData2(N,I,I,s)` saves N simulations in a comparable form equal to `callpcod` in a $14 \times 5 \times N$-double-matrix $SimData$, `getInfoSimData2(SimData)` returns mean and variance of respective data points in 14×5-matrices, M and V. Moreover `PlotSim2(SimData)` provides illustration of $SimData, M,\ and\ V$ compared to the Table 2 in the wet case.

[7] If no input values are given, the user will be asked to provide them in this order.
[8] s can be considered the number of colonies created by inoculation.

	wet					dry				
Col. size	day1	day 2	day 3	day 4	total	day1	day 2	day 3	day 4	total
1-10	115	91	49	19	274	168	29	40	37	274
11-20	21	13	22	3	59	23	13	8	19	63
21-30	4	10	2	0	16	4	6	3	5	18
31-40	1	4	2	1	8	1	3	5	1	10
41-50	3	2	2	3	10	3	3	1	1	8
51-60	0	1	4	2	7	0	0	2	2	4
61-70	0	0	4	3	7	0	0	3	1	4
71-80	0	1	2	0	3	0	2	0	2	4
81-90	0	0	0	1	1	0	1	1	0	2
91-100	2	2	0	0	4	2	0	2	0	4
100-500	0	4	7	2	13	0	1	2	0	3
500-1000	1	0	0	2	3	1	1	0	0	2
1000+	1	1	1	0	3	1	0	0	0	1
total	148	129	95	36	408	203	59	67	68	397

Table 2: Experimental data by Dulla & Lindow

6. PHYLLOSIM - a spatial approach

The following provides a brief overview over the model by *Annemieke van der Wal et. al.*.
In this spatial model, a 2D-grid consists of 100×100 elements with periodic boundary conditions each representing 100 μm^2 of the cuticle. The reproduction dynamics are based on nutrient consumption and biomass gain. Depending on the water constellation nutrients like sugar photosynthates appear on the leaf surface via diffusion and are consumed by the individual cells in each time step.

Parameter	Symbol	Value	Unit	Reference	Notes
System size		$1 * 10^6$	μm^2		
Grid element size		100	μm^2		
Maximum growth rate	μ_{max}	$1.11 * 10^{-4}$	s^{-1}	[14]	Doubling time 1.7 h
Substrate affinity constant	K_s	0.3	g m^{-3}	[14]	
Concentration of sugars in apoplast	C_{apo}	18	g m^{-3}	[45]	
Permeability of the cuticle	P	$2.78 * 10^{-10}$	m s^{-1}	[15]	
Fructose requirement per cell doubling	f	$3.0 * 10^{-13}$	g	[14]	
Initial conditions					
Average of the number of bacterial cells per 1 mm^2 domain	N_0	10			
Average normalized biomass of each bacterial cell	B_0	1.5			
Equations					
1) $A = \pi * sin^2\alpha * (3*V/(\pi * (2-3*cos\alpha+cos^3\alpha)))^{2/3}$			m^2	[15]	A = contact area of water drop α = contact angle
2) $C_{sink(t+\Delta t)} = (V * C_{sink(t)} + \Delta t\ (F_{(t)} - U_{(t)}))/\ V$			g m^{-3}	[15]	V = volume of water drop
3) $F_{(t+\Delta t)} = A * P * (C_{apo} - C_{sink(t+\Delta t)})$			g s^{-1}	[46]	*F* = Flow of sugar from the apoplast to the sink (water drop)
4) $G_{i(t+\Delta t)} = B_{i(t)} * \mu_{max} * C_{sink(t+\Delta t)}/(C_{sink(t+\Delta t)}+K_s)$			s^{-1}		Monod kinetics G_i = growth of biomass of bacterium i B_i = normalized biomass of bacterium i (dimensionless) C_{sink} = Concentration of sugars in the water drop
5) $U_{(t+\Delta t)} = f * \Sigma(G_{i(t+\Delta t)})$			g s^{-1}	[14]	U = uptake of sugars summed over all bacteria in a water drop
6) $B_{i(t+\Delta t)} = B_{i(t)} + \Delta t * G_{i(t+\Delta t)}$					Δt = time step, 60 s
Rules					

[1]If $B_i >= 2$, the bacterial cell divides. The biomass is split equally between the parent and daughter cell.
doi:10.1371/journal.pone.0075633.t001

Figure 3: Exert from the PHYLLOSIM-article ([3], Table1)

Each cell also has, besides its Biomass B, an identification number *id*. If the biomass B exceeds a value of 2, the cell splits and biomass is divided equally, the daughter cell receives the same *id*. The individual cell volume $V = B * 1\mu m^3$, presuming a height of $1\mu m$, lead to an area coverage $B * 1\mu m^2$ of a grid element. If the number of cells in one grid element surpasses 100 cells newly created bacterial cells will assort themselves in one of the adjacent grid elements randomly. Fig.3 provides an impression of the dynamics of PHYLLOSIM.
The authors first tried to explain and simulate the dispersion of water on the leaf by regarding different scenarios, the 'null' model where a continuous water film covers the phyllosphere and the 'patchy water' models where four water drops, capable of varying in size, reside upon the leaf surface. As the observations they refered to were made under circumstances of 100% relative humidity they neglected effects like vaporisation.
Upon receiving unsatisfactory simulation results which did not seem to differ greatly they introduced migration dynamics into the model. The cell *id* system was modified so migrated cells were assigned a new id providing information about the resident colony.[9] This yielded qualitatively better results and like *Pérez-Velazquez et. al.* they concluded the detachment and relocation (and possibly re-attachment) to take a primary role in observed patterns.

source: [3]

[9] At the inoculation all inhabitants of a colony bear the same *id*.

Part III.
Analysis and Improvement of the stochastic Model

In this part the stochastic approach is examined more closely and improvements concerning assumptions will be attempted and partially implemented. Criticism and improvement proposals are based on the article [3] and my own perception and conclusions. The focus lies on migration rate, capacity conditions and inoculation circumstances.

7. Migration Rate

```
1   while T(j)<Tmax
2
3       bN=max([r*N(j)-r*N(j)^2/(K),0]);     %birTh probabiliTies model 1
4       dN=d*N(j);                           %deaTh probabiliTies model 1
5       iN=I;                          %immigration probabiliTies model 1
6       lambda=bN+iN+dN;
7       T(j+1)=T(j)-log(rand)/lambda; %add expoNeTial disTribuTed iNTer-eveNT iNTerval of Time
8       a(j)=T(j+1)-T(j);
9       ra=rand;
10      if ra < bN/lambda;
11          N(j+1)=N(j)+1;       %iT was a birTh
12      elseif ra < (bN+dN)/lambda;
13         N(j+1)=N(j)-1;       %iT was a deaTh
14      else
15          N(j+1)=N(j)-1;    %iT was an Immigration
16        aux(j+1)=j;
17      end;
18
19      if N(j+1)==0             %Check To see if N(j)==0
20          j=j+1;               %fiNish ouT program
21          T(j+1)=Tmax;
22          N(j+1)=0;
23      end; %if
24      j=j+1;                   %iNcremeNT couNTer
25
26  end; %while
```

Figure 4: while-loop of `pcod1.m`, apart from denotation similiar to `pcod2.m`

Although not explicitly mentioned in the article [1] of *Pérez-Velázquez et. al.*, in their implementation only initial colonies (created by inoculation) are capable of migration but the migration rate is not modified for new colonies. `pcod1.m` and `pcod2.m` are equivalent, only their output is treated differently. In line 16 of Fig.4 the array `aux` holding required data for colony foundations simply remains unused for second-generation colonies.
Therefore all initial colonies have a constant migration rate I while daughter colonies have an effective migration rate of 0. Filial generation colonies still have migration events but in case of a migration event no new colonies will be found. The migration rate I then works as an additional death rate for non-initial colonies. I didn't find any biological reason why the second-generation could not bring up colonies themselves.

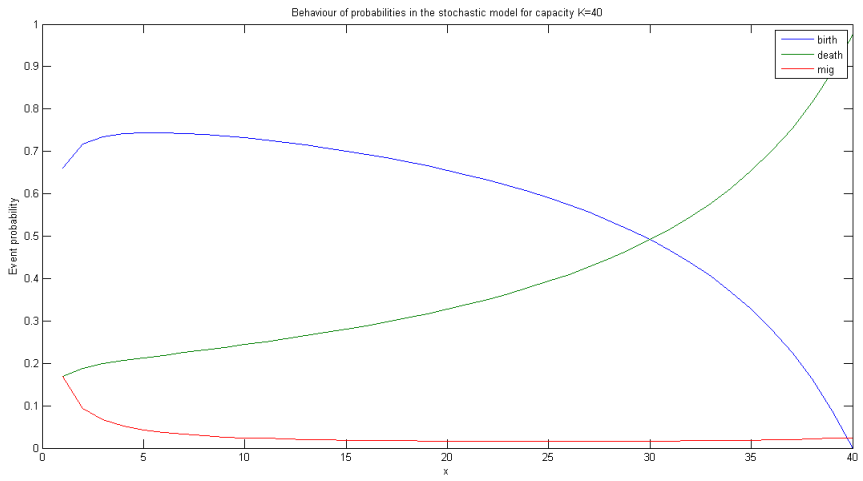

Figure 5: Examplary Migration probability h(N)

The constant migration rate poses another controversal subject because the migration probability decreases for high population values. The probability of a migration event (else case, line 14) happening is $P(Next\ event\ is\ migration) = P(\frac{bN+dN}{R_{tot}} \leq ra \leq 1) = P(ra \leq \frac{iN}{R_{tot}}) = \frac{iN}{R_{tot}} = \frac{I}{rx-\frac{r}{K}x^2+dx+I} =: h(x)$ with population number[10] x, $ra \sim unif(0,1)$ and R_{tot} as the total event rate (line 6). Due to logistic growth x does not exceed the capacity K, so $x \in [0, K]$. For $r = 0.4, d = 0.1$ and $I = 0.1$, chosen like in the stochastic simulation:[11]

$$\frac{\partial}{\partial x}h(x) = -\frac{I}{R_{tot}^2}(r - 2\frac{r}{K}x + d) = \begin{cases} < 0 & x \in [1, \frac{5}{8}K) \\ = 0 & x = \frac{5}{8}K \\ > 0 & x \in (\frac{5}{8}K, K] \end{cases}$$

The boundaries yield $h(1) = \frac{1}{6-\frac{4}{K}}$ and $h(K) = \frac{1}{K+1}$ leading to $h(1) \underset{K\to\infty}{\to} \frac{1}{6}$ and $h(K) \underset{K\to\infty}{\to} 0$. For $K > 4$ is $h(1) > h(K)$ hence for higher capacities K the migration chance reaches its maximum for colonies with only one resident, for example in the case of newly created colonies. Fig.5 gives an impression of the development of $h(N)$(red line) and birth(blue) and death(green) probabilities for a capacity $K = 40$.

It is debatable if this is qualitatively suitable. The migration probability should be increasing with augmenting population x towards K as the capacity displays the availability of resources and intraspecific stress. Aspects like intraspecific cooperation cannot justify such extreme behaviour of the migration probability. Regarding the situation in Fig.5 one can observe that $\frac{h(1)}{h(40)} = 6.9492$. It seems highly unrealistic that it is almost seven times more likely for a solitary resident which may just have arrived via migration or inoculation to migrate again than the case that in a brimmed colony one bacterial cell leaves to settle elsewhere. This boundary limit relation grows more unrealistic with increasing K. For colonies restricted by $K = 1000$ which do appear in the data the relation $\frac{h(1)}{h(1000)} = 166.9446$ becomes inapplicable.
Another problem is the assumption that migrating cells always create new colonies. Although cells are able to sense their environment, to say apprehend the capacity K and the population number x, spatial aspects like already occupied adjacent residential areas are entirely neglected. Especially for longer observation durations regional occupancy of potential colony sites plays a relevant role as it naturally will be getting more diffult for migrating cells to find new uninhabited sites as travel distances increase and settling sites get rarer. However, introducing dynamics following such aspects in a stochastic model will most certainly prove to be a highly difficult task.

[10] To avoid confusion with the different rates bN, dN, iN the population number is x instead of N.

[11] $r - 2\frac{r}{K}x + d \overset{!}{=} 0 \Leftrightarrow x = \frac{r+d}{2r}K$

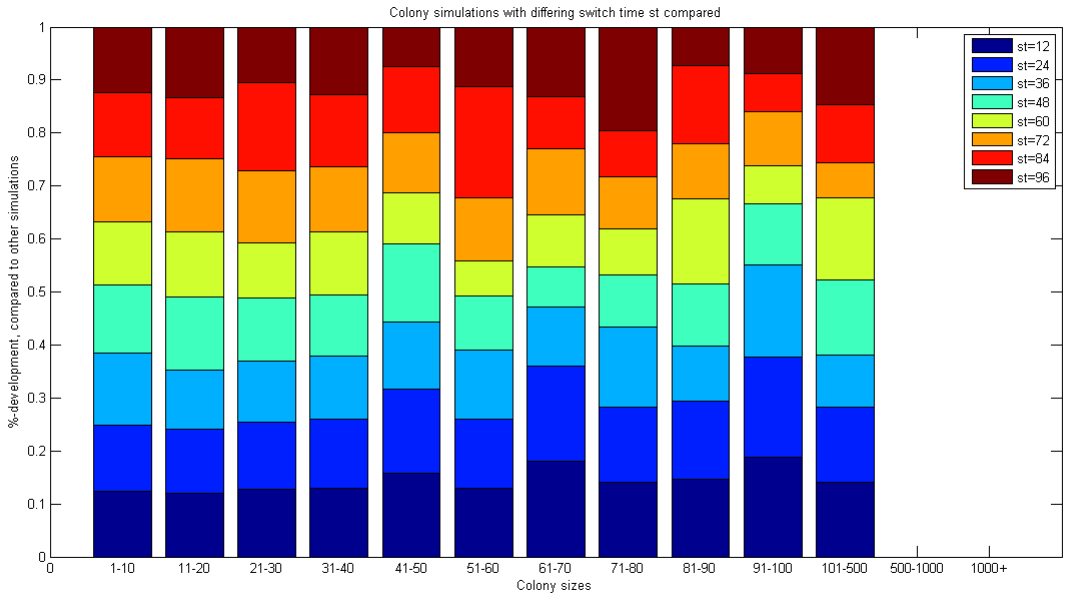

Figure 6: Simulations with different switchtimes compared, $I_1 = 0.1$ and $I_2 = 0.05$

st denotes switchtime, each colour represents the mean of 25 simulations with differing *st*

If one tries to amend the two migration rates for parental and filial generations setting the rates in dependence of time seems most reasonable. `LBD_multiple_wanted_2MigRates_time.m` produces a simulation including an alternation of the migration rate at a specified time. Setting the second migration rate lower than the first seems apparent as resources and available residential sites get exhausted. Assuming that the influence the switch time point should bear will increase and decrease with the difference between the two migration rates seems plausible.
But Fig.6 shows that the switch time between migration rates $I_1 = 0.1$ and $I_2 = 0.05$ did not have a visible impact.
Another idea is to enable migration to colonies but in a stochastic model. Realisation attempts will probably be fruitless because, beside their difficulty, independence assumptions cannot be made anymore, making analysis even trickier. I'm confident that spatial aspects like local regions devoid of unoccupied settling sites are the main reason for bacterial cells to reattach. But in a stochastic model one must rely on other colonies' status like capacaties and the current number of residents and cannot take, for example, locally limited sensing abilities of bacteria or the interspace between colonies into account.
Changing the constant migration rate could bear better results. A linear migration rate $iN := Ix$, with $R_{tot} = r + d + I - \frac{r}{K}x$, is a good reasonable assumption:

From the previous section we get $\frac{\partial}{\partial x}h(x) = \frac{iN}{R_{tot}^2}\frac{\partial}{\partial x}R_{tot} = \frac{Ir}{K(r+d+I-\frac{r}{K}x)^2} \geq 0$ for $x \in [1, K]$. The migration probability $h(x)$ is strictly increasing, the lower boundary $h(1) = \frac{I}{r(1-\frac{1}{K})+d+I}$ is the minimum and the upper boundary $h(K) = \frac{I}{I+d}$ the maximum. Also, the relation $\frac{h(1)}{h(K)} = \frac{d+I}{r(1-\frac{1}{K})+d+I} \leq 1$ appears more sensible.
For standard parameters, denoted in 1, Fig.7 shows the behaviour of the probabilities for different migration rate alternation attempts if the capacity $K = 40$. In the linear case migration and death rate are equal for the chosen simulation parameters, so iN covers dN. The linear alternation seems like a promising way to improve the model, the qualitative behaviour looks consequential.[12]

[12] I regard the constant migration as inapplicable. If not stated otherwise the linear migration rate will be used

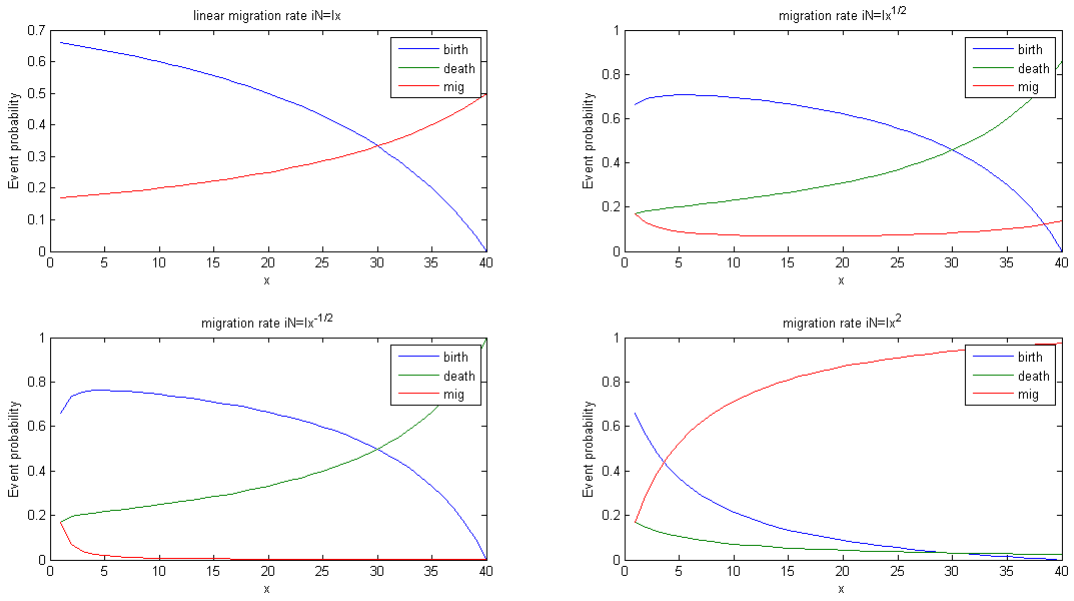

Figure 7: Examplary plots of migration probability h(N) for different migration rates

8. Capacity

In the stochastic model a priori assumptions concerning the capacities are made. The authors of the PHYLLOSIM article [3] criticized this and stated that the model by *Pérez-Velázquez et.al.* is imposing a pattern rather than explaining.
I also consider the dynamics of a log-normal-distribution problematic and share their opinion. Log-normally distributed capacities $K_i \sim ln\mathcal{N}(\mu, \sigma)$, in the model $\mu = 1.4511$ and $\sigma = 1.7025$, represent the availability of nutrients hence providing a tool to measure and compare the attractiveness of different colony sites. Obviously high capacity resorts should be occupied first but in the model the capacities K_i are determined irrespective of variation in time. If the migration chances are extremely low or non-existent later on the extent of this effect is severely reduced aswell but, beside reasons mentioned in the previous two sections, since low-populated colonies are still found within the data in Table 2 at any observation point a strongly diminishing or vanishing migration rate appears unsuitable. Generally the dynamics introducing capacities in the system are very stiff, for example, considering fusion of colonies looks infeasible.
Examining the data one can observe in both cases, wet and dry, a colony containing more than 1000 cells on the first day which vanished by the fourth day. This leads to the problem that one colony is hardly able to reach a population of 1000 starting with an initial size of one cell within $24h$. Running `simWaiting_multiple_runs(0.4,0.1,0.1,2500,20000,1000)` provided interesting results:
From 1000 colonies with a capacity 2500 each only 478 were able to reach a population of 1000 cells within 20000 event steps. The elapsed time until one these 478 colonies first reached 1000 inhabitants averaged at $41.9874h$ and the fastest colony required $30.8351h$. Thus, none was able to reach the required 1000 cells within one day although they had very high capacities. If $K_i \sim ln\mathcal{N}(1.4511, 1.7025)$ then $P(K_i > 1000) = 1 - P(K_i \leq 1000) = 1 - 0.999324884334486 = 0.0006751156655141477 \approx 0.07\%$, so it is very unlikely for a colony to be assigned a capacity greater than 1000. In addition to a probably insufficient population growth speed reaching this threshold within one day seems unlikely.

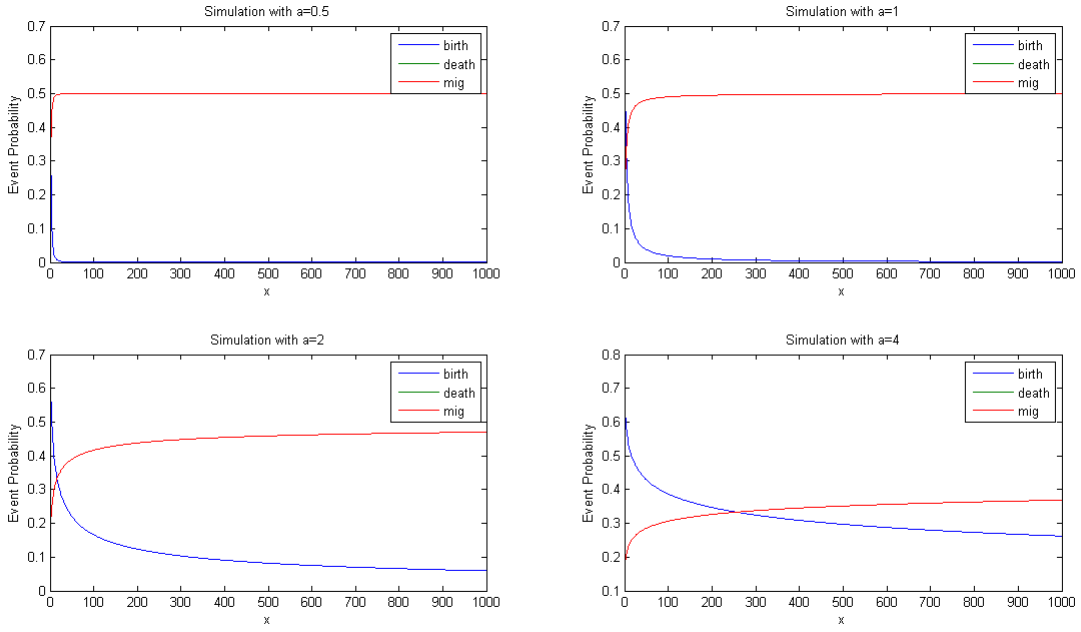

Figure 8: Probability behaviour for different a-values, time-independant, $r = 0.4,\ d = 0.1$ and $I = 0.1$

Capacity-bounded colonies aren't flexible enough to match the data. The biomass dynamics of the PHYLLOSIM-model avoid capacity boundaries instead it respects the room each cell requires and ultimately can handle the spatial extentension increasing colonies demand up to possible fusions. The growth of colonies is based on individual nutrient consumption and availability so resource competing (including spatial effects) should eventually lead to saturation.
However introducing biomass dynamics in a stochastic model will hardly work without making radical assumptions. A possibilty is to abolish the capacity and replace the logistic birth rate $bN = rx(1 - \frac{x}{K})$ by $\widetilde{bN} := rx(\frac{1}{\sqrt[a]{x}}) = rx^{\frac{a-1}{a}}$, where the positive parameter $a \in \mathbb{R}$ regulates saturation. Higher values for a result in higher birth rates, so instead of a high capacity a greater a-value stands for better environmental conditions.
The birth probability $b(x) := \frac{\widetilde{bN}}{\widetilde{\lambda}}$, $\widetilde{\lambda} = \widetilde{bN} + dN + iN$ is strictly decreasing as $\frac{\partial}{\partial x} b(x) = \frac{-\frac{r}{a} x^{\frac{a-1}{a}} (d+I)}{\widetilde{\lambda}^2} \leq 0$ for all $x > 0$ and $\lim\limits_{x \to \infty} b(x) = \lim\limits_{x \to \infty} \frac{r}{r + (d+I) x^{\frac{a+1}{a}}} = 0$.
Fig.8 plots some examplary a-values, note that the migration rate was already assumed to be linear like suggested in the previous section.
As long as the birth rate $\widetilde{bN}$ is greater than the death and migration rate dN and iN it appears intuitive that the colony size will tend to grow and vice versa.

$$\widetilde{bN} \geq dN + iN \Leftrightarrow rx^{\frac{a-1}{a}} \geq (d+I)x \Leftrightarrow x^{-\frac{1}{a}} \geq \frac{d+I}{r} \Leftrightarrow x \leq \left(\frac{r}{d+I}\right)^a \tag{13}$$

Table 3: Development of $X_{brd} = 2^a$

a	1	2	4	6	8	10
X_{brd}	2	4	16	64	256	1024

The threshold $X_{brd} := \left(\frac{r}{d+I}\right)^a$ obtained from Eq.13 with given $r = 0.4, d = 0.1, I = 0.1$ takes the values given in Table 3. Obviously X_{brd} is an indicator for the target value population sizes will eventually fluctuate around.

One could replace the parameter a by a time-dependant monotone decreasing function to further administer to nutrient depletion or residential sites occupancy.

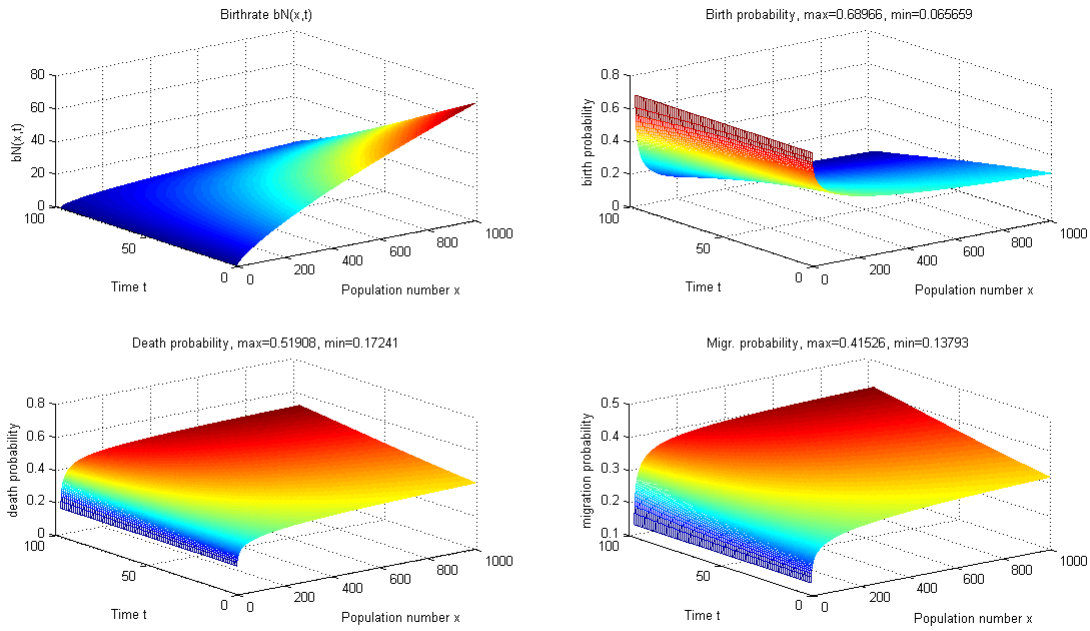

Figure 9: Qualitative behaviour of birth rate and event probablities,
$\beta = 4,\ \alpha = 2,\ r = 0.4,\ d = 0.1$ and $I = 0.08$

For example, having an observation duration from t_0 to t_{end}, setting $a(t) = \beta - (\beta - \alpha)(\frac{t-t_0}{t_{end}-t_0})$, with $\beta \geq \alpha \geq 0,\ \beta \neq 0,\ t \in [t_0, t_{end}]$ provides a linear decrease from $a(t_0) = \beta$ to $a(t_{end}) = \alpha$. Fig.9 displays the impact on birthrate and event probabilities for mentioned a.

9. Inoculation

As already mentioned in the section seven it turns out to be difficult for colonies to reach 1000 cells within one day starting with a single resident. Increasing the initial population sizes for colonies at the inoculation point will probably help adjusting the model to the data.
Fig.10 and Fig.11 give an impression how initial sizes affect the model with logistic and modified time-independant birth rate, bN and $\widetilde{bN}$ respectively, (both with linear migration rate). The logistic birth rate will probabably fluctuate around $X^* := K(1 - \frac{d+I}{r})$ which leads to $a \approx 10.3$ if $X_{brd} = X^*$ and parameters chosen like in Fig.10.[13]

Table 4: Birth probabilites for bN and $\widetilde{bN}$

$Pop.x$	bN	$\widetilde{bN}$
1200	50.98%	50.10%
1000	54.55%	50.54%
750	58.33%	51.24%

Table 4 and Fig.11 demonstrate that the birth probability ascertained by $\widetilde{bN}$ with $a = 10.3$ approaches values in the proximate periphery of 0.5 much faster than the capacity-bounded. Hence $\widetilde{bN}$ eventually requires additional time to reach the population size $x = X_{brd}$ as x-increasing and x-decreasing events are almost equiprobable quite early. Trying to simulate the required time for $\widetilde{bN}$ with $a = 10.3$ like in Fig.10 did not procure usable results as the simulated populations do not often reach values of 1000 or greater in 20,000 steps.(mean (maximal population reached) was between 900 and 950, see folder 'PartThree-Inoculation')

[13] $X^* = 2500(1 - \frac{0.1+0.1}{0.4}) = 1250$ and $1250 = 2^a \Leftrightarrow a = 10.2877$

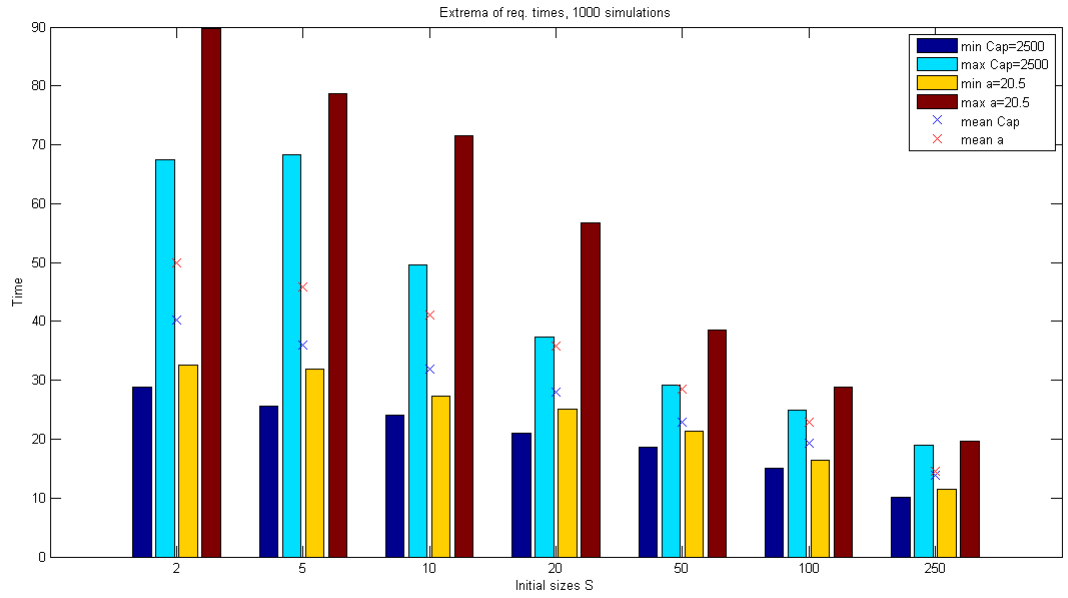

Figure 10: Simulation (20,000 events) of required time to reach a population size of 1000 for different initial sizes with capacity $K = 2500$, $a = 20.5$, $r = 0.4$, $d = 0.1$ and $I = 0.1$

To let $\widetilde{bN}$ decrease less rapidly it could be modified by for example drawing the root, to say $\widetilde{bN}^* := rx\sqrt{x^{-\frac{1}{a}}} = rx^{\frac{2a-1}{2a}}$, but this is obviously equivalent to simulations produced with $\widetilde{bN}$ with doubled parameter a as displayed in Fig.10.

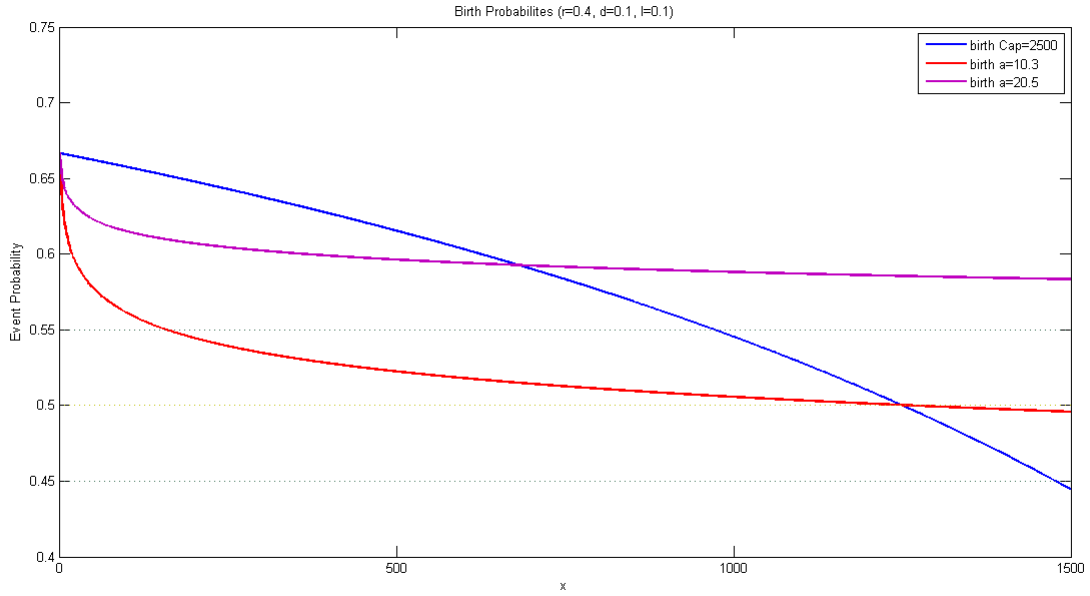

Figure 11: Behaviour of birth probabilities for differing birth rates

The PHYLLOSIM-model uses a Poisson-distribution to spread the cells upon the cuticle. For the stochastic model one could follow similiar dynamics by introducing this distribution for the initial population.
As the Poisson-distribution is a discrete distribution it is suitable for generating initial sizes. If the initial sizes $S \sim Pois(\lambda)$ with probability mass function $f_\lambda(k) = P(S = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ the mean and variance are λ.

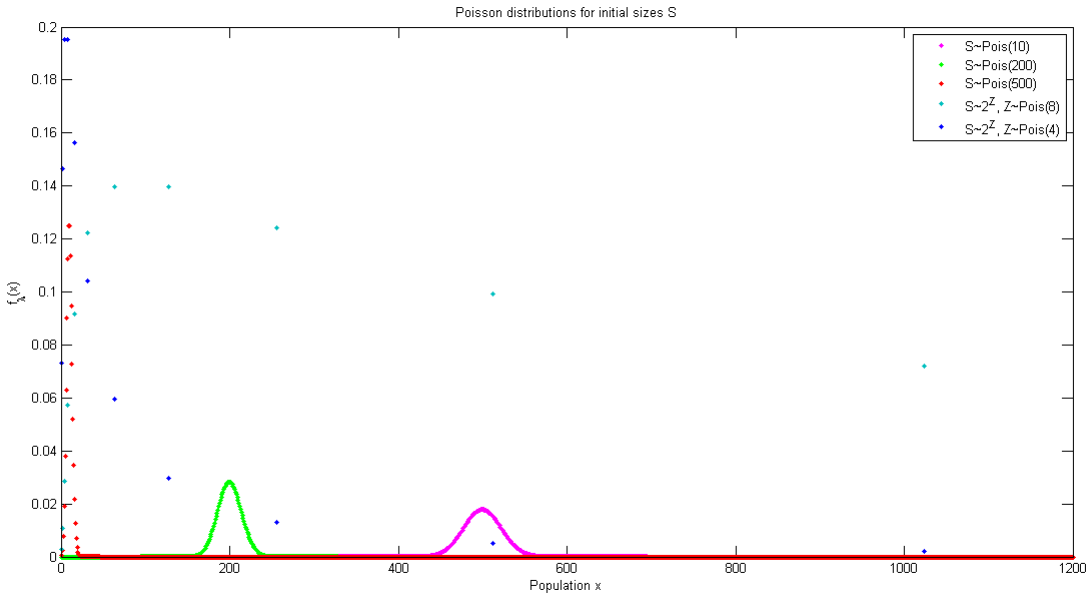

Figure 12: Examplary initial size distributions

Although it is stated in both articles that over 50% of cells are in colonies larger than 1000 it is still difficult to estimate λ with the given data, table 2, which is not very accurate.
It seems apparent regarding Fig.12 that $S \sim Pois(\lambda)$, which assigns a probability to every population size procures an unsuitable centered distribution. Having only rough data to rely on, assuming $S \sim 2^Z$, $Z \sim Pois(\lambda)$ seems practical. A constant $c \in (0, 1)$ can modify $S \sim \lceil c2^Z \rceil$, $Z \sim Pois(\lambda)$ further to adjust the inital sizes.

Part IV.
Conclusion

10. Summary and Overview

I suggest to modify the stochastic model for simulation time from t_0 to t_{end} in the following way:

Birth rate	$bN = rx^{\frac{a(t)-1}{a(t)}}$	$r \in \mathbb{R}_+$
	$a(t) = \beta - (\beta - \alpha)\frac{t-t_0}{t_{end}-t_0}$	$\beta, \alpha \in \mathbb{R}_+,\ \beta > \alpha \geq 0$
Death rate	$dN = dx$	$d \in \mathbb{R}_+$
Migr. rate	$iN = Ix$	$I \in \mathbb{R}_+$
Init. sizes	$S \sim 2^Z$	$Z \sim Pois(\lambda),\ \lambda \in \mathbb{R}_+$

Setting $\beta \sim Pois(\xi)$, $\xi \in \mathbb{R}_+$ and $\alpha \sim \beta X$ where X shall be distributed on the interval $(0, 1)$ provides good dynamics. Besides $unif(0, 1)$ the Beta-distribution $beta(\nu, \eta)$ with *pdf* $f_{\nu,\eta}(x) = \frac{1}{B(\nu,\eta)}x^{\nu-1}(1-x)^{\eta-1}I_{[0,1](x)}$, *beta*-function $B(\nu, \eta) := \frac{\Gamma(\nu)\Gamma(\eta)}{\Gamma(\nu+\eta)}$ and indicator function $I_{[0,1]}$ seems apt. $\Gamma(n) := (n-1)!$ denotes the *gamma*-function. Fig.13 gives an impression of normalized $\alpha \sim \beta(\nu, \eta)$ compared to the uniform-distribution.

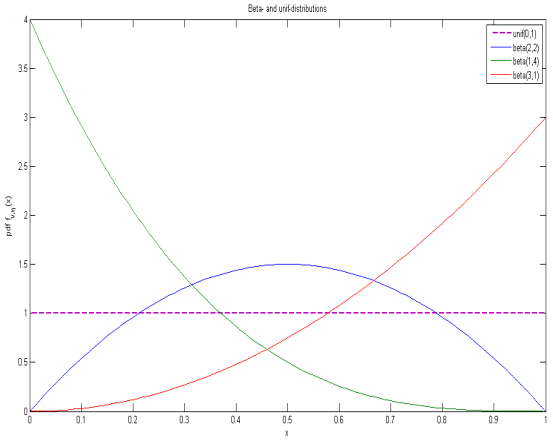

Figure 13: Visualisation of $beta(\nu, \eta)$ and $unif(0, 1)$

Overall 8 parameters appear (time boundaries t_0 and t_{end} excluded): $r, d, I, \lambda, N, \xi, \nu$,and η. The function `Model_multiple_wanted(r,d,I,lambda,N,xi,nu,eta,t0,t_end,wantedTimes)` will simulate this model suggestion accordingly and is capable of procuring data comparable to the data, Table 2 if $wantedTimes = [24\ 48\ 72\ 96]$. Some examplary simulations were done, see folder 'PartFour-Summary and Overview'.

Examining the stochastic model by *Pérez-Velázquez et. al.* has revealed some inconsistence with the related article and inappropriate assumptions or dynamics. Especially the fact that the model cannot pay heed to spatial aspects is a major issue as migration on the phyllosphere heavily depends on regional circumstances, as Table 2 regarding dry and wet conditions suggests. I believe an approach with a stochastic model to be rather unsuccessful. Suggestions made in this thesis may aid in matching the simulation to the data but in the long term a spatial model like PHYLLOSIM is probably a superior foundation.
Additional and more precise data would also greatly help; as of now it is difficult to even be certain of good results in particular concerning big colony sizes. It is interesting that the author teams of both models concluded that migration is essential although they examined the issue from partly vastly differing angles, I am convinced of this aswell.
I reckon that the stochastic model is not suitable for simulating *Pseudomonas Syringae* bacteria dispersions upon the leaf surface.

11. Future Work

Originally the bachelor thesis' objective was to estimate parameters for the stochastic model by *Pérez-Velázquez et. al.* and efforts have been made in this direction. Great parts of the entire code were created under this prospect thereby creating a decent basis for paramter estimation or further model development. The attached CD contains all code that was produced for this thesis in the folder 'Bachelor'.
Tables 5 and 6 briefly explain some functions that could turn out useful. The explanations are not precise, so it is advised to have a closer look at the code and its comments to gain a better understanding.
If one tries to develop the model I presented further I would suggest to work on the migration dynamics, specifically on reattachment. My belief is that this grows more important with greater simulation durations but I reckon it highly difficult to find suitable dynamics, even regardless of run-time efficiency. The introduction of the time-dependant, birth rate modifying function $a(t) = \beta - (\beta - \alpha)\frac{t-t_0}{t_e-t_0}$ gave the impression that simulations required much longer so I suspect reattachment dynamics to deteriorate run-times even more.
Performing parameter estiamation with the provided code should be relatively easy and quite prompt. Nevertheless due to the absence of accurate and numerous data I recommend to try to acquire more data, even of other bacterial cultures, to assure a good foundation first.

parameter	$\in$	meaning/application
r	$\mathbb{R}_+$	birth rate
d	$\mathbb{R}_+$	death rate
I, I_1, I_2	$\mathbb{R}_+$	migration rate
N	$\mathbb{N}_+$	number of simulations
S	$\mathbb{N}_+$	number of initial colonies
N_E	$\mathbb{N}_+$	number of events
x_0, x_0^*	$\mathbb{N}_+$	initial size (vector)
K, K^*	$\mathbb{N}_+$	capacity (vector)
a	$\mathbb{R}_+$	$\widetilde{bN} = rx^{\frac{a-1}{a}}$
t_0, t_s, t_e	$\mathbb{R}_+$	times
st	$\mathbb{R}_+$	migration switch time
st	$\mathbb{R}_+$	migration switch time
wT	$\mathbb{R}_+$	wanted times (vector)
λ	$\mathbb{R}_+$	$x_0 \sim Pois(\lambda)$
ξ	$\mathbb{R}_+$	$\beta \sim Pois(\xi)$
(ν, η)	$\mathbb{R}_+$	$\alpha \sim \beta beta(\nu, \eta)$
α, β	$\mathbb{R}_+$	$a(t) = \beta - (\beta - \alpha)\frac{t-t_0}{t_e-t_0}$
$SimData$	$\mathbb{N}_+$	$14 \times 5 \times N$-matrix

Table 5: Input parameters and meaning

Function name	Input	Explanation
`CapacityGenerator`	ν, σ	generates $log\mathcal{N}(\nu, \sigma)$-distributed rounded values/capacities
`DataOverview`	x	returns 14×5-matrix comparable to data. x is output of function including `wanted` with $wT = [24\ 48\ 72\ 96]$
`PlotSim2`	$SimData$	produces 2 plots, comparing with experimental data (Wet case)
`getInfoSimData2`	$SimData$	returns two 14×5-matrices containing mean/variance of entries of $SimData$
`getSimulationData2`	N, I_1, I_2, S	saves N simulations of `LBD_multiple_wanted_2MigRates` as $SimData$
`getSimulationDataTime`	N, I_1, I_2, st, S	saves N simulations of `LBD_multiple_wanted_2MigRates_time` as $SimData$
`getSimulationData_model`	$r, d, I, \lambda, S, \xi, \nu, \eta, N$	saves N simulations of `LBD_multiple_wanted_2MigRates_model` as $SimData$
`LBD_single_wanted`	r, d, I, K, x_0, t_0, wT	simulates one colony, capacity bounded, returns migration times
`LBD_single_wanted_2MigRates_time`	$r, d, I_1, I_2, st, K, x_0, t_0, wT$	simulates one colony, capacity bounded, migration rate changes at sT, returns migration times
`LBD_multiple_wanted`	$r, d, I, K^*, x_0^*, t_0, wT$	simulates multiple colonies, uses `LBD_single_wanted`
`LBD_multiple_wanted_2MigRates`	$r, d, I_1, I_2, K^*, x_0^*, t_0, wT$	simulates multiple colonies, uses `LBD_single_wanted`
`LBD_multiple_wanted_2MigRates_time`	r, d, I_1, I_2, st	simulates multiple colonies, uses `LBD_single_wanted_2MigRates_time`
`Model_single_wanted`	$r, d, I, \beta, \alpha, x_0, t_s, t_0, t_e, wT$	simulates single colony, returns migration times, t_s= colony start time, t_0/t_e=sim. start /end time
`Model_multiple_wanted`	$r, d, I, \lambda, N, \xi, \nu, \eta, t_0, t_e, wT$	simulates multiple colonies, uses `Model_single_wanted`
`sim_waitingtimes`	r, d, I, K, N_E	simulates N_E events for single colony, capacity-bounded
`sim_waitingtimes_init`	r, d, I, K, x_0, N_E	simulates N_E events for single colony, initial size x_0, capacity-bounded
`sim_waitingtimes_BirthMod_init`	r, d, I, a, x_0, N_E	simulates N_E events for single colony, initial size x_0
`simWaiting_multiple_runs`	r, d, I, K, N_E, N	N multiple events, returns time pop. reaches 1000, uses `sim_waitingtimes`
`simWaiting_multiple_runs_init`	r, d, I, K, x_0, N_E, N	N events, returns time pop. reaches 1000, uses `sim_waitingtimes_init`
`simWaiting_multiple_birthMod`	r, d, I, a, x_0, N_E, N	N events, returns time pop. reaches 1000, uses `sim_waitingtimes_BirthMod_init`
`simWaiting_multiple_birthMod_max`	r, d, I, a, x_0, N_E, N	multiple events, returns time pop. reaches max, uses `sim_waitingtimes_BirthMod_init`

Table 6: Outline of MATlab-Functions

Acknowledgements

I want to thank my supervisors Dr. Judith Pérez-Velázquez and Dr. Christina Kuttler for their help and advice which they generously offered. Also, I want to give my thanks to Jan Hasenauer who assisted me in the early stages of the thesis.

References

[1] J.Pérez-Velázquez et. al. : *Stochastic modeling of Pseudomonas syringae growth in the phyllosphere.* MBS 7238 2.June 2012

[2] Eric Renshaw: *Modelling Biological Populations in Space and Time*, Cambridge University Press, 1991, ISBN 0-521-30388-5

[3] Annemieke van der Wal et. al. : *Explaining Bacterial Dispersion on Leaf Surfaces with an Individual-Based Model (PHYLLOSIM).* `http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0075633` (Stand 12.3.2014)

[4] Jan Hasenauer and Fabian Theis: *Parameter inference for stochastic and deterministic dynamic biological processes.* `lectureParamEst.pdf` on attached disc

[5] Daniel T. Gillespie : *Exact Stochastic Simualtion of Coupled Chemical Reactions.* The Journal of Physical Chemistry, Vol. 81, No. 25, 1977

Attachment:

CD with MATlab programs