A Stochastic Formulation of Bacterial Attachment in a Spatially Explicit Model of Cellulolytic Biofilm Formation

by

Yousef Rohanizadegan

A Thesis presented to The University of Guelph

In partial fulfilment of requirements for the degree of Master of Science in Mathematics

Guelph, Ontario, Canada

© Yousef Rohanizadegan, December, 2018
ABSTRACT

A STOCHASTIC FORMULATION OF BACTERIAL ATTACHMENT
IN A SPATIALLY EXPLICIT MODEL OF CELLULOYTIC BIOFILM FORMATION

Yousef RohanizadeganAdvisor:
University of Guelph, 2018Dr. Hermann J. Eberl

We propose a mathematical framework for introducing random attachments of bacterial cells in continuum models of biofilms. Our approach deploys the formalism of stochastic differential equations to model an auxiliary stochastic process that drives impulses of bacterial cells. We especially apply the proposed framework to a spatially explicit model of cellulolytic biofilm formation, which comprises a coupled PDE-ODE system that governs the bacterial biomass and the carbon substrate. The model equations are discretized in space by a standard finite volume method and temporally integrated by explicit numerical schemes. We explore some computational and programming solutions for improving the speed and efficiency of simulations and preventing instability issues. Our numerical simulations reproduce the specific features of cellulolytic biofilms with cell attachments. Grid refinement studies show convergence for the expected values of spatially integrated biomass density and carbon concentration. We also examine the sensitivity of random attachments to a few model parameters.
I would like to first and foremost thank Dr. Eberl for his thoughtful supervision, intellectually stimulating conversations, patience and kind considerations. My introduction to the world of applied mathematics has definitely helped me develop much needed computational skills and hone my analytical and problem solving ones as well. I am very grateful to the members of my advisory committee, Dr. Willms and Dr. Fryxell, for their patience and constructive comments on the present work. I also express my gratitude to Dr. Sonner (Radboud University) for her reviewing of our stochastic model and giving us advice on the technical details of stochastic differential equations.

I wish to acknowledge with gratitude the generous help I have received in the form of graduate research assistantships, graduate teaching assistantships, university scholarships and external scholarships from Dr. Eberl, Department of Mathematics & Statistics, College of Physical & Engineering Science and the Government of Ontario.

Finally, I would like to thank my family, especially my mother for her support from afar and also express my gratitude to my sister and my brother-in-law here in Canada for their technical help and kind advices.

I wish to dedicate this thesis to the memory of my late father, Dr. Rahim Rohanizadegan, for he always took pride in my scholarly pursuits.
# Table of Contents

Abstract ii

Acknowledgements iii

Table of Contents iv

List of Tables vi

List of Figures vii

1 Introduction 1

1.1 Biofilms . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1
1.2 Attachment Phenomenon . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3
1.3 Modeling Objective and Application . . . . . . . . . . . . . . . . . . . . . . 4
1.4 Cellulolytic Biofilms . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
1.5 Background of Cell Attachment Modeling . . . . . . . . . . . . . . . . . . . 8
1.6 A New Approach . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 10
1.7 Overview of the Thesis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12

2 Mathematical Model 13

2.1 A Spatially Explicit Continuum Model . . . . . . . . . . . . . . . . . . . . . 13
2.1.1 Model Assumptions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
2.1.2 Governing Equations . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
2.1.3 Spatial Discretization . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
2.1.4 Regularization of the Model Equations . . . . . . . . . . . . . . . . . . 20
2.2 A Spatially Explicit Model with Random Cell Attachments . . . . . . . . . . 21
2.2.1 Stochastic Differential Equations . . . . . . . . . . . . . . . . . . . . . . 22
2.2.2 Mathematical Formulation of Stochastic Cell Attachment . . . . . . . . 24
2.2.3 Non-negativity and Boundedness in the Regularized Problem . . . . . . 29

3 Computational Implementation 33

3.1 Numerical Methods . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
3.2 Parallelization . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 36
3.3 Random Number Generation ........................................... 41
3.4 Random Number Generation in Parallel .......................... 44
3.5 Numerical Stability and Time Step Size ............................ 46

4 Simulation Results and Computational Analysis .................... 49
  4.1 2D Simulations .......................................................... 49
  4.1.1 Case 1: Initially Clean Domain ................................. 51
  4.1.2 Case 2: Initially Colonized Domain ........................... 58
  4.1.3 Case 3: Spatial Variation in Random Attachments .......... 62
  4.2 Grid Refinement Studies .............................................. 65
    4.2.1 Refinement and Convergence of Solution Components .... 69
    4.2.2 Refinement and Stochasticity ................................. 81
  4.3 Parameter Sensitivity Analysis ..................................... 84

5 Discussion ........................................................................... 90

6 Conclusion ........................................................................... 96
  6.1 Lessons Learned .......................................................... 96
  6.2 Future Work ............................................................... 96

Bibliography ........................................................................ 98
List of Tables

3.1 Hardware and software specifications. . . . . . . . . . . . . . . . . . . . . . . . . 37

4.1 Model parameters used for 2D simulations. . . . . . . . . . . . . . . . . . . . 50
4.2 Model parameters that are different in 1D simulations. . . . . . . . . . . . . . 65
List of Figures

1.1 Schematic diagram of a biofilm community with filament-like colonies . . . . . 3
1.2 Schematic diagram showing the six stages of cellulolytic biofilm formation . . 6
1.3 Distribution of \( C. \text{obsidiansis} \) cells on a cellulose chad . . . . . . . . . . . . . 7
1.4 Crater formed by \( C. \text{thermocellum} \) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7
1.5 An example of simulated PDE-ODE system from [12] . . . . . . . . . . . . . . . 9

2.1 A representation of the sample path of a standard Wiener process, \( W_t \) . . . 23
2.2 The impulse function, \( f(\phi) \) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

3.1 The speedup as a function of number of threads with PRNG turned off . . . . . 38
3.2 The speedup as a function of number of threads for serial PRNG . . . . . . . . 39
3.3 The speedup as a function of number of threads for parallel PRNG . . . . . . . 40

4.1 Simulation of (2.38) with initially \( M = 0 \) everywhere and \( \psi = 0.08 \) . . 53
4.2 Simulation of (2.38) with initially \( M = 0 \) everywhere and \( \psi = 0.04 \) . . 55
4.3 Sample paths with different \( \psi \) values for one component of the AF, \( \phi \), vs. time 57
4.4 Change in the fraction of cells within the support of \( f(\phi) \) vs. time . . . . . 57
4.5 Simulation of (2.38) with initially \( M \neq 0 \) only in the center and \( \psi = 0.08 \) . 59
4.6 Simulation of (2.38) with initially \( M \neq 0 \) only in the center and \( \psi = 0.04 \) . 61
4.7 Simulation of (2.38) with spatial variation in random attachments . . . . . 64
4.8 1D Simulation of (2.38) with size of random attachments \( \frac{1}{16} \) of \( L \) . . . . . 66
4.9 1D Simulation of (2.38) with size of random attachments \( \frac{1}{128} \) of \( L \) . . . . . 67
4.10 1D Simulation of (2.38) with size of random attachments equal to a cell . . . . 68
4.11 Total biomass and carbon vs. time for several grid sizes. Attachment size: \( \frac{1}{16} L \) 71
4.12 Total biomass and carbon vs. time for several grid sizes. Attachment size: \( \frac{1}{128} L \) 71
4.13 Averaged total biomass and carbon vs. time. Attachment size: \( \frac{1}{16} L \) . . . . . 72
4.14 Averaged total biomass and carbon vs. time. Attachment size: \( \frac{1}{128} L \) . . . . . 73
4.15 Averaged total biomass and carbon with errors. Attachment size: \( \frac{1}{16} L \) . . . . 74
4.16 Averaged total biomass and carbon with errors. Attachment size: \( \frac{1}{128} L \) . . . . . 75
4.17 Standard error of the mean for biomass and carbon. Attachment size: \( \frac{1}{16} L \) . 76
4.18 Standard error of the mean for biomass and carbon. Attachment size: \( \frac{1}{128} L \) . 76
4.19 2-norm convergence of spatially integrated solutions. Consecutive grid sizes. . . 78
4.20 2-norm convergence of spatially integrated solutions. Coarser grids wrt finest 79
4.21 1-norm of the standard deviations of spatially integrated solutions . . . . . . 80
4.22 Averaged total biomass and carbon for different grid sizes. Attachment size: $\frac{L}{N}$ 82
4.23 Averaged total biomass and carbon with errors. Attachment size: $\frac{L}{N}$ . . . . . 83
4.24 Standard error of the mean for biomass and carbon. Attachment size: $\frac{L}{N}$ . . 84
4.25 Sensitivity of the mean values of the total biomass and carbon to $\psi$ . . . . . 86
4.26 Sensitivity of the mean values of the total biomass and carbon to $a \& b$ . . . 87
4.27 Sensitivity of the mean values of the total biomass and carbon to $\sigma$ . . . . . 88
5.1 The 1-norm of standard deviation in a 1D setting vs. the number of samples 93
5.2 The 1-norm of standard deviation in a 2D setting vs. the number of samples 94
Chapter 1

Introduction

This chapter provides the introductory material for this thesis. We start by giving a basic description of biofilm formation and biofilm properties. We lay out the objective of the thesis, which is the study of the stochastic attachment of free cells in the environment that contains a bacterial biofilm community. We especially look into this topic in the context of cellulolytic biofilms. We examine a few modeling strategies that either have reproduced such random attachments or can be potentially used to realize them and discuss their advantages and disadvantages. Finally, we elaborate on our own approach and its accomplishments before we close by a summary of the remaining chapters at the end.

1.1 Biofilms

Bacterial biofilms are mainly formed by accumulation of microorganisms on solid surfaces inside a fluid environment. These microorganisms are embedded in a self-secreted matrix known as the extracellular polymeric substance (EPS) that provides an adhesion mechanism to the surface as well as structural stability to the biofilms. The EPS is assumed to be mostly made of polysaccharides, but it may also contain proteins and nucleic acids. The microorganisms (or cells) are sessile and protected against environmental hazards inside the EPS [22, 36].

Like many other microbial communities biofilms can be either detrimental or beneficial. For example, biofilms can cause infectious disease and, due to their structure, are hard to combat using antibiotics [22]; biofilm formation on the surfaces of medical devices and medical implants can harm their functionality and cause device failure and infection [55]; dental plaque, a prime cause of tooth decay, is a biofilm formation on the surface of teeth.
that shows low susceptibility to antimicrobial agents due to its biofilm structure [42]; biofilm growth inside oil pipelines, due to injection of seawater for secondary oil recovery, leads to problems such as corrosion and reduction of flow [50]. Aside from all the aforementioned adverse effects, biofilms are also deployed for their benefits in some areas of environmental engineering. For instance, biofilms are grown on wastewater filters to eliminate biological pathogens and other suspended microorganisms; contaminated soil can be treated by biofilms where contaminants are biodegraded as a product of biofilm growth [58, 33]; biofilms are utilized for generating electricity from a variety of organic materials in microbial fuel cells [46].

The solid surface that supports the biofilm structure is called the substratum. Some substrata are impenetrable inert surfaces such as rocks, sand and plastic. However, substrata can also be reactive and play an essential role in the biofilm development *i.e.* supply the biofilm with nutrients for growth (see Section 1.4); or degrade as a side effect of biofilm growth as in the corrosion of oil pipelines and the dissolution of enamel due to dental plaque [58].

Biofilms need nutrients for their growth. These nutrients, also known as substrates, are either delivered to the biofilm by the substratum or absorbed from the surrounding environment that transports substrates through mechanisms such as diffusion and convection. Depending on the availability, rate of delivery and local concentrations of substrates, the growth of biofilm may be slow and limited in some locations and fast and ample in other ones. Assuming that the availability of substrates is not a growth limiting factor, the metabolic growth rate of the biofilm will determine the generation of biological by-products and the uptake of substrates [58].

A biofilm community often consists of many separated colonies, which might merge as they expand to form bigger colonies. From a macroscopic point of view, a biofilm community possesses a homogeneous structure. In other words, it appears as a thin film of organic material that covers a particular substratum. However, such structural form would not be as descriptive when one transitions to mesoscopic scale (10 μm-1mm in this case). At these scales the heterogeneity of a biofilm community manifests itself. Factors such as maximum cell density, specific growth rate and substrate concentration determine the level of heterogeneity of a biofilm community. For instance, in situations where the substrate concentration and hence the nutrient delivery rates are higher, a thick homogeneous biolayer forms where most colonies have merged, whereas the deficiency of substrate concentration results in a patchy filament-like heterogeneous colony formation [54] (see Figure 1.1).
1.2 Attachment Phenomenon

The biofilm formation on a substratum is generally initiated by the physical attachment of free floating bacteria/cells from the surrounding aqueous phase that is then followed by adhesion\(^1\). One possible source of free floating cells is biomass\(^2\) detachment as a result of shear forces exerted by the fluid on the biofilm [54]. The possibility of adhesion and formation of a thriving bacterial colony at a certain position on the substratum can be influenced by a number of factors including [2]

- Environment: temperature, bacterial concentration, chemical treatment, etc.,
- Bacterial characteristics: hydrophobicity, surface charge, appendages, etc.,
- Substratum properties: chemical composition, roughness, physical configuration, etc.

Therefore, the initial attachment and the concomitant adhesion may depend on numerous model parameters when modeling these in a mathematical framework is the primary goal. Consequently, because of the sheer number of parameters required to characterize the attachment and adhesion of cells, the mathematical formulation of these phenomena could turn out to be too complex. Hence, one possible way to simplify the mathematical modeling of the

\(^1\)Formation of a firm physiochemical bond.

\(^2\)For modeling purposes, everything inside the biofilm structure is treated as one solid lump and labeled “biomass” [58].
attachment phenomenon is to think of it as an emergent random occurrence where the effect of all the deterministic underlying mechanisms, which are controlled by the parameters, are lumped together and treated as a single stochastic process. In other words, we view the attachment phenomenon as a random behavior of the floating cells in the aqueous phase for which we need not concern ourselves with the underlying deterministic processes. Whatever these processes are, we treat them as a stochastic driving force that causes a single attachment to take place arbitrarily at any point in space and time. As soon as such attachment of a cell occurs, a new colony can start developing.

1.3 Modeling Objective and Application

The mathematical modeling of stochastic cell attachment and the possible effects it may have on the growth and temporal evolution of biofilms is the main focus of the present thesis. The overall goal is to introduce a straightforward mathematical framework for modeling the random cell attachment and to perform numerical simulations by which the model can be tested and numerically analyzed. Since modeling the random attachments using an established theoretical framework is of paramount importance to us, we have deployed the formalism of Itô stochastic differential equations [47] in our approach (for details see Section 2.2.1), which provides both the theory and the numerical methods [29] we require.

Barring computational costs, the modeling framework established here could potentially be applied to biofilms of different types and morphologies and in as many spatial dimensions as required.

For testing and analysis purposes, we have applied the developed stochastic attachment formalism to the special case of cellulolytic biofilms. These types of biofilms are mainly used in the biofuel production industry and have their own characteristics (see Section 1.4). In experiments conducted on these biofilms, it has been observed that cells in the aqueous phase randomly attach to the available sites on the substratum and start new colonies [11, 57]. This attachment process is a discrete phenomenon and is not well-understood and characterized in terms of adhesion driving factors discussed in Section 1.2. Our mathematical approach to model stochastic attachment of cells in the context of cellulolytic biofilm formation is largely based on these observations and premises.
1.4 Cellulolytic Biofilms

Biofuels are one of the alternatives for use in the transportation sector in place of fossil fuels. Among the benefits of using biofuels are the abundant renewable resources and the net zero carbon dioxide release into the atmosphere. The production of biofuels from food products, such as sugar, starch, or fats and oils, are not as sustainable as their production from inedible resources, *e.g.* lignocellulose\(^3\), due to the ever increasing demand for food [5]. In the US, the annual production of cellulosic ethanol from lignocellulosic resources such as woody biomass, switchgrass, and crop waste was estimated at 53 billion gallons in 2016 [31].

The commercial production of cellulosic ethanol in large quantities is deemed achievable for lignocellulosic biomass. However, a few difficulties arise with this type of feedstock: biomass recalcitrance, which is associated with the resistance of cell walls to enzymatic deconstruction; the heterogeneity of the substrate, which arises from the complex structure of lignocellulosic feedstock [35]. As a result, the biomass needs to be treated and prepared before ethanol can be extracted. In consolidated bioprocessing (CBP) preparation and production stages, namely enzyme production, polysaccharide hydrolysis and sugar fermentation, are combined in one single operation [35, 49]. CBP requires microorganisms that are highly efficient at producing cellulolytic enzymes and at fermenting resultant sugars; at the same time they should produce small amounts of other byproducts and be able to live in high levels of ethanol [49].

Strains of bacteria used in CBP, such as *Clostridium thermocellum* and *Caldicellulosiruptor obsidiansis*, consume a cellulosic substratum and form cellulolytic\(^4\) biofilm colonies, which are distributed spatially and evolve temporally [11, 44, 57]. These kinds of biofilms tend to produce little amounts of EPS. They attach directly to the substratum and form monolayer biofilms [11, 57]. The regular biofilm formation produces bacterial layers that grow into the aqueous phase and absorb growth nutrients from it whereas the cellulolytic biofilms do not form such structures and instead consume and degrade the cellulosic substratum. This mode of operation carves crater-like structures in the substratum that are called inverse colonies [57]. The cellulolytic biofilm formation and growth progresses through a number of stages (see Figure 1.2)

1. Initial cell attachment to the substratum,

\(^3\)Lignocellulose is composed of carbohydrate polymers (cellulose, hemicellulose), and an aromatic polymer (lignin).

\(^4\)Hydrolyzing or having the capacity to hydrolyze cellulose.
2. Cell growth and division,

3. Inverted colony formation,

4. Formation of crater-like depressions due to substrate consumption and substratum degradation,

5. Radial growth of the depressions and their eventual coalescence,

6. Formation of uniform-thickness biofilm.

Figure 1.2: Schematic diagram showing the six stages of cellulolytic biofilm formation: 1) initial cell attachment, 2) cell growth and division, 3) inverted colony formation, 4) emergence of crater-like depression, 5) growth and merging of craters, 6) uniform biofilm formation [57].

As mentioned in Section 1.2, cell detachment from a biofilm community occurs via different mechanisms, e.g. due to shear forces exerted by the surrounding fluid. The detachment phenomenon is extant in the case of cellulolytic biofilms as well, and free cells do exist inside the aqueous phase (as a consequence of detachment) as observed in experimental settings [11, 57].

The different stages of spatiotemporal evolution of biofilm formed by C. obsidiensis on cellulose chads were captured by a confocal laser scanning microscope [57] and are demonstrated in Figure 1.3. The details of a crater-like structure and its depth are provided in Figure 1.4 also from [57].
Figure 1.3: Distribution of *C. obsidiansis* cells on a cellulose chad after a) 0h, b) 8h, c) 16h, d) 24h, e) 44h, f) 48h, g) 56h and h) 68h. Radius of craters $\approx 40\mu m$ [57].

Figure 1.4: Crater formed by *C. thermocellum*. a) top view, b) cross-sectional view. Length scale = 10$\mu m$ [57].
1.5 Background of Cell Attachment Modeling

The mathematical modeling of stochastic attachment of cells has been addressed in an ad-hoc manner by modifying a deterministic model of cellulolytic biofilms [12]. Here, the deterministic part deals with the formation and temporal evolution of cellulolytic biofilms and is described by a continuum model, namely a highly nonlinear degenerate coupled system of a partial differential equation (PDE) and an ordinary differential equation (ODE). The biomass production and mechanics is governed by the PDE and the uptake of the substrate is governed by the ODE. The ad-hoc inclusion of stochastic cell attachment is basically implemented by the addition of an impulse function to the biomass PDE: if a certain quantity \( Q \) is greater than a uniform random number \( U[0,1] \), the impulse function becomes non-zero and a fixed amount of biomass is added to the PDE; otherwise the value of the impulse function is zero and no biomass is added. The quantity, \( Q \), is defined by using a couple of probabilistic arguments: an attachment position is less probable to be occupied if it already accommodates some biomass (M) and if it is deficient in substrate concentration (C) i.e. \( Q \propto (1 - M) \times C \). By formulating the attachment phenomenon this way, the proposed continuum model in [12] has reproduced the qualitative behavior of cellulolytic biofilms with cell attachment. One can easily confirm this claim by examining the graphical visualization (see Figure 1.5) of the numerical simulation for this model and compare it with experimental observations shown in Figure 1.3. Although this approach is successful in generating qualitative results that resemble the experimental observations, it is not based on the mathematical framework of stochastic processes [12]. That is, it is not supported by a stochastic theory, namely the theory of stochastic differential equations. Since our goal is to introduce a modeling framework that can be supported by established stochastic theories, especially the theory of continuous stochastic processes, we shall concern ourselves with modeling strategies that fulfill this requirement and are potentially able to reproduce the same qualitative results obtained by the ad-hoc approach.

One possible path to modeling the random attachment of cells is to formalize it as random noise. For instance, one could add a term containing white noise to the biomass PDE in [12]. This will turn it into a stochastic PDE (SPDE), which requires the rigorous analysis of SPDEs for computational purposes [37]. However, this modeling approach is not a fully viable solution and a few undesirable side effects are associated with it, which include negative

---

5Uniform distribution over \([0,1]\).
6White noise is a sequence of independent random variables with zero means and equal finite variances.
biomass values — especially at locations with zero initial biomass — and uncontrollable rate and magnitude of cell attachment. These are due to the nature of white noise, or more specifically additive white noise\(^7\), which can assume both positive and negative values with arbitrary magnitudes. Even if the biomass PDE is first discretized in space and then turned into a system of stochastic differential equations (SDEs) by adding white noise, the above problems persist and one cannot easily do away with them.

One could argue that by making certain assumptions and “tweaking” the noise term, more specifically making its coefficient depend on state variables, the negative values could be excluded from the solution. For instance, in the case of systems of semi-linear parabolic SPDEs it has been shown that if the deterministic and stochastic parts meet certain conditions, the non-negativity of the solution is guaranteed [8]; or in case of systems of (ordinary) SDEs, even necessary and sufficient conditions for the non-negativity and/or boundedness of solutions are known [9]. But of course adjusting the equations to satisfy these conditions requires justifiable mathematical/physical arguments. Another possible remedy is directly modifying the numerical schemes that estimate the solution by including reflecting or absorbing barriers: reflecting barriers send negative values to their corresponding positive counterparts and absorbing barriers swap negative values with zeros (\textit{e.g.} [19, 26, 53] have deployed these to avoid negative values in the stochastic version of their models). But then certain questions

\(^{7}\)A type of noise term that does not depend on the state variables, for example, it does not vanish when \(M = 0\).
arise with regard to the relation of the numerical simulations and the mathematical model and whether the simulations are valid tests of the actual model’s behavior.

Assuming the negativity issue is eliminated by a legitimate approach, the uncontrollability of attachment makes the simple addition of white noise problematic. All positive values of the noise term correspond to attachment events (detachment events for negative noise values) for which the amount of the attached bacteria is an arbitrary value. This behavior has been corroborated by our numerical simulations as they have shown that the high frequency of attachment instances and the randomly varying amount of attaching bacteria in each instance do not qualitatively replicate the experimental observations e.g. the prominent features of cellulolytic biofilms shown in Figure 1.3 and Figure 1.4. This problem follows from the mathematical properties of the added white noise. Since the noise is primarily realized by sampling the normal distribution in the numerical simulations of SPDEs and SDEs, the high frequency of attachment events and their arbitrary magnitude are natural outcomes.

All the issues detailed so far follow from the simple addition of white noise to the studied PDE-ODE system. Of course it would be very straightforward, theoretically tenable and elegant, in terms of a modeling approach, if such simple formulation could give rise to a model that replicates the qualitative aspects of the experimental observations and has solutions that are non-negative and bounded. However as it turns out, a more controlled and calculated approach is needed here. We have addressed all these issues in Section 1.6 where we lay out our attempt on introducing stochasticity into the model in [12].

1.6 A New Approach

The starting point in our approach to model the random attachment of cells is to conceive a mechanism that is not characterized by the factors mentioned in Section 1.2 but rather by a local stochastic property of the substratum and the environment. We call this local stochastic property, the attachment factor (AF)\(^8\). It is assumed to be a continuous stochastic process that can be roughly interpreted as a particular local background noise in the system. We also assume that the AF is driven by a local Itô SDE whose stochastic part is proportional to the local values of state variables (see Section 2.2.2). The SDE’s deterministic part is assumed to be zero so that there are no contributions to the temporal evolution of the AF from

---

\(^8\)Later when we prepare the model for numerical calculations by introducing a spatial grid, each grid cell will have its own AF. Thus, it is local in this sense.
any deterministic sources. We have chosen this particular treatment of the deterministic part over any other since any choice other than a zero-valued term must entail legitimate assumptions and justifications.

The random cell attachments are discrete events, which are to be included in the modeling framework developed in [12]. Our approach treats these random events as impulsive phenomena. Therefore, we consider an impulse function\(^9\) that provides the means for introducing the stochastic cell attachment to arbitrary locations on the substratum and also allows controlling different aspects of the attachment. Most importantly, the issue with negative solutions can be readily addressed since we have the freedom to choose a function with suitable properties as the candidate for the impulse function: a function that always outputs positive values will not add negative biomass especially to locations where there is no biomass. Furthermore, the effective domain of the chosen impulse function can be adjusted by its characterizing parameters. For instance, one may manipulate the frequency of attachments by careful exploitation of the function parameters that affect its effective domain. Also, the amount of attaching bacteria in a single event can be controlled by the coefficient of the impulse function.

The numerical simulations of our approach have visually reproduced the experimental observations depicted in Figure 1.3 and Figure 1.4. Adopting this particular modeling strategy has proven to be successful in addressing the issues we laid out in Section 1.5. We have also introduced the framework of stochastic differential equations into the model, which has a strong theoretical foundation and well-established numerical methods.

In order to obtain numerical solutions for the new model that includes the random attachments, the governing biomass PDE is discretized spatially by a uniform grid in the same manner performed in [12]. The discretization converts the PDE into a system of ODEs where each ODE governs the biomass dynamics locally in a specific grid cell. The stochasticity in each cell is simulated via the corresponding locally-defined Itô SDE, which governs the AF. The conversion of the PDE to a system of ODEs has the added benefit of allowing us to use well-known ODE and SDE solvers for each cell. We have used explicit Euler and Euler-Maruyama methods for the numerical integration of the model ODEs and SDEs, respectively. We have made this choice in order to capture the effect of stochasticity in the model. The smaller time steps needed for the stability of explicit methods, will make the simulation of the AF’s sample path more precise. Hence, capturing the stochastic attachment of bacteria, which is our primary goal, will be accomplished more accurately. The

\(^9\)The local value of the AF operates as the independent variable of the impulse function.
larger time steps achieved by utilizing implicit methods make them unsuitable to accurately capture the stochastic variations in the sample path of simulated SDEs; thus, there will be lesser chance to capture the stochasticity in the sample solutions of the ODEs.

1.7 Overview of the Thesis

In Chapter 2 we provide a detailed description of the original model introduced and developed in [12]. The mathematical description of our approach to model the random attachment of cells from the aqueous phase is the other major part of this chapter.

Chapter 3 details the computational realization of the numerical simulations. We used a few strategies to speed up our code and reduce the simulation times, which we have elaborated here.

The simulation results are all contained in Chapter 4. We provide visualizations for 2D simulations. The convergence analysis and sensitivity analysis of this model can be found in this chapter as well.

Chapter 5 is composed of some discussion on the limits and encountered difficulties as well as some more elaboration on details of our specific approach to modeling, computational realization and analysis of results.

Lastly, we close the thesis by concluding remarks and possible future work in Chapter 6.
Chapter 2
Mathematical Model

This chapter begins with a detailed introduction to the cellulolytic biofilm PDE-ODE model developed in [12], which is then followed by a description of the discretization method utilized to transform the PDE-ODE system into a system of ODEs. We also explain how a regularized version of the system of ODEs will converge to the continuous PDE-ODE model. The final part of this chapter details our approach to introducing stochastic attachment of cells. Here we elaborate on the modeling of this phenomenon and how the SDE formalism is embedded into the deterministic version of the model.

2.1 A Spatially Explicit Continuum Model

The deterministic differential-equation models are established on the assumption that the spatiotemporal transformation of an ensemble of individuals arises from the average behavior of its constituents [12]. These types of models allow a more concise and straightforward formulation of the underlying mechanisms, which can readily lend themselves to the modeling of larger scale processes e.g. reactor scale. The main purpose of the present model is to provide a framework where the qualitative features of cellulolytic biofilms can be simulated. By applying some standard visualization methods to the simulation results, one can replicate the experimental observations [11, 57] and through careful considerations possibly propose new mechanisms that could explain the special features of such biofilm communities.

In the following sections, we lay out the model assumptions and the governing equations. We show how the model is prepared for numerical simulations. We also remark on the well-posedness of the solutions in the case of a regularized version of the model and how the regularized model is linked to the main model.
2.1.1 Model Assumptions

The proposed PDE-ODE system [12] is basically an adaptation of a previous model [14] to the case of cellulolytic biofilms. Here, the aqueous phase does not explicitly appear in the model formulation. The model assumptions are the following:

1. The local capacity per unit area/volume of the substratum for accommodating bacterial biomass is finite.

2. The only consumed substrate for biomass production is carbon. Since this substrate is part of the chemical composition of the substratum, it is deemed to be immobile i.e. it does not diffuse.

3. The spatial distribution and movement of biomass is only dependent on the available space and nutrient. As a consequence, the biofilm does not distinctly expand into neighboring positions if the local amount of available space and nutrient is adequate for biomass development.

4. The uptake of the substrate and the growth of the biomass are driven by standard saturation kinetics.

5. The cell loss from the biofilm occurs in two ways: cell death and cell detachment. The cell loss rate is constant; thus, the amount of lost biomass is proportional to the biomass density.

6. There are no attachments of cells from the aqueous phase.

2.1.2 Governing Equations

The model comprises a PDE that governs the spatiotemporal evolution of the biomass density, $M$ and an ODE that describes the temporal evolution of carbon concentration, $C$, due to consumption by the biomass. Namely, the equations are

$$
\partial_t M = \nabla \cdot (D(M) \nabla M) + F(C)M, \quad \text{for } t \in \mathbb{R}^+, \ x \in \Omega, \quad (2.1)
$$

$$
\partial_t C = -G(C)M, \quad \text{for } t \in \mathbb{R}^+, \ x \in \Omega. \quad (2.2)
$$

Here $M := M(t, x)$, $C := C(t, x)$, $D(M)$ is the diffusion coefficient, $F(C)$ is the net biomass growth rate and $G(C)$ is the uptake rate of the carbon substrate. The spatial domain is
\( \Omega \subset \mathbb{R}^d, d = 1, 2, 3 \), which is assumed to be rectangular in the two-dimensional case. Due to high computational costs, the model is only studied for \( d = 1 \) and \( d = 2 \) cases.

The function \( D(M) \), is the density dependent diffusion coefficient [14, 28] given by

\[
D(M) = \delta \frac{M^\alpha}{(M_\infty - M)^\beta}, \quad \delta > 0, \quad \alpha \geq 1, \quad \beta \geq 1, \quad (2.3)
\]

where \( \delta \) is the motility coefficient and \( M_\infty \) is the maximum biomass density. Here \( \alpha \) and \( \beta \) determine the magnitude of the driving force for different diffusion modes that come into effect depending on the local density of biomass.

The proposed diffusion coefficient exhibits two distinct behaviors: (i) if the local biomass density approaches the maximum carrying capacity, the extra produced biomass moves to the neighboring positions on the substratum (in accordance with model assumption 1); (ii) if sufficient space is available to the newly produced biomass, it will be accommodated locally (in accordance with model assumption 3). These behaviors emerge from the underlying non-linear modes of diffusion. For \( 0 < M \ll M_\infty \), the diffusion coefficient can be approximated as a degenerate power law, i.e. it vanishes for \( M = 0 \), which is similar to the porous medium equation: \( D(M) \approx \delta M^\alpha \). It should be noted that the degenerate power law creates an expanding biofilm colony that has a sharp distinct front; that is, in mathematical terminology, initial data with compact support generates solutions with compact support [13]. Here, the value of \( \alpha \) determines how much the local biomass density can grow before a noticeable expansion to neighboring regions takes place. When \( 0 \ll M < M_\infty \), the diffusion coefficient approaches singularity and super diffusion becomes dominant, which prevents the biomass density to approach its maximum value [13]: \( D(M) \approx (M_\infty - M)^{-\beta} \). Here, the value of \( \beta \) controls the strength of this effect. It should be emphasized that these nonlinear diffusion modes work in tandem to produce the finite-speed expansion (due to porous medium non-linearity) and the local volume filling (due to super diffusion nonlinearity) of biomass and neither can cause these effects alone [12].

The net biomass growth rate consists of two terms. The first term is the true biomass growth rate and the second term is the cell loss rate. Since model assumption 2 deems carbon as the sole nutrient consumed by the biomass for growth, the true growth rate is assumed to be proportional to the carbon concentration in the form of Monod equation [45] (model assumption 4). The cell loss, per model assumption 5, occurs through two different mechanisms: cell death and cell lost to the aqueous phase. Both of these mechanisms are combined together in a constant term called the cell loss rate. Hence, the net growth rate is
given by
\[ F(C) = \mu \frac{C}{\kappa + C} - \lambda, \]  
(2.4)
where \( \mu \) is the maximum specific growth rate and \( \lambda \) is the cell loss rate.

The uptake rate of the carbon substrate is formulated by standard saturation kinetics in the form of Monod equation (model assumption 4) as follows
\[ G(C) = \Upsilon \frac{C}{\kappa + C}, \]  
(2.5)
where \( \kappa \) is the half saturation concentration and \( \Upsilon \) is the maximum consumption rate.

The substrate uptake rate in the given form is the only driving factor in Eq. (2.2). There is no transport mechanism contributing to the substrate concentration since carbon is assumed to be immobilized in the substratum (per model assumption 2). This is where the current model deviates from the original framework in [14]. The uptake rate attains a constant maximum value when substrate is locally abundant for consumption. In the case of limited local amounts of substrate, the uptake rate becomes proportional to the available carbon concentration.

The biomass is assumed to be confined to the domain \( \Omega \). For this to be the case, the homogeneous Neumann boundary condition is imposed, namely
\[ \partial_n M = 0, \quad \text{for} \quad t \in \mathbb{R}^+, \quad x \in \partial \Omega. \]  
(2.6)
where \( \partial_n \) denotes the normal derivative and \( \partial \) in \( \partial \Omega \) designates the boundary of \( \Omega \). The biomass density initial condition is
\[ M(0, x) = M_0(x), \]  
(2.7)
with the assumption that \( M_0(x) > 0 \) for a few small patches. We have \( M_0(x) \equiv 0 \) in the rest of \( \Omega \).

The initial condition for the carbon concentration is
\[ C(0, x) = C_0(x), \quad \text{for} \quad x \in \Omega. \]  
(2.8)
where we consider \( C_0(x) > 0 \). Although \( C_0(x) \) is position dependent, in the simulations of this model it is assumed to be at its maximal value, \( C_\infty \), for the whole domain, which is in agreement with experiments conducted on homogeneous paper chads as the cellulosic
substrata [10, 11, 57].

The existence and uniqueness of solutions has been proven for the systems of equations where Eq. (2.1) is coupled to one or two substrate equations that, unlike Eq. (2.2), in addition to the reaction term undergo Fickian diffusion as well [16, 17]. We assume the well-posedness of the PDE-ODE system here can follow from the established existence-uniqueness proof, applied to the case of non-diffusing substrate.

Diffusion-reaction equations of the type given in Eq. (2.1) with the specific diffusion coefficient (2.3), have few associated analytical treatments available in the literature. As a result they are mainly studied via numerical methods. In preparation for computer simulations, the model is cast into the non-dimensional form by scaling $x, t, M$ and $C$ as follows

\[
\begin{align*}
\tilde{x}_i &= \frac{x_i}{L} \quad \text{for} \quad i \leq d, \\
\tilde{t} &= \mu t, \\
\tilde{M} &= \frac{M}{M_\infty}, \\
\tilde{C} &= \frac{C}{C_\infty},
\end{align*}
\]

where $L$ is a characteristic length of $\Omega$; for instance, it can be chosen as the length of the longest side of a rectangular domain. Here $1/\mu$, the characteristic time of biomass growth, is the time scale of the dimensionless problem. The dimensionless parameters are given by

\[
\begin{align*}
\tilde{\kappa} := \frac{\kappa}{C_\infty}, \\
\tilde{\Upsilon} := \frac{M_\infty \Upsilon}{C_\infty \mu}, \\
\tilde{\lambda} := \lambda \mu, \\
\tilde{\delta} := \frac{\delta M_\infty^{(\alpha-\beta)}}{\mu L^2}.
\end{align*}
\]

Consequently, the model equations in their non-dimensional form are

\[
\begin{align*}
\partial_t \tilde{M} &= \tilde{\nabla} \cdot (\tilde{D}(\tilde{M}) \tilde{\nabla} \tilde{M}) + \tilde{F}(\tilde{C}) \tilde{M}, \\
\partial_t \tilde{C} &= -\tilde{G}(\tilde{C}) \tilde{M},
\end{align*}
\]
where

\[ \tilde{D}(\tilde{M}) = \tilde{\delta} \frac{\tilde{M}^\alpha}{(1 - \tilde{M})^\beta}, \]  
(2.13)

\[ \tilde{F}(\tilde{C}) = \frac{\tilde{C}}{\tilde{\kappa} + \tilde{C}} - \tilde{\lambda}, \]  
(2.14)

\[ \tilde{G}(\tilde{C}) = \tilde{\Upsilon} \frac{\tilde{C}}{\tilde{\kappa} + \tilde{C}}. \]  
(2.15)

The initial data in the non-dimensional form for a homogeneous cellulosic substratum are

\[ \tilde{M}(0, \tilde{x}) = \tilde{M}_0(\tilde{x}), \quad \tilde{C}(0, \tilde{x}) = 1. \]  
(2.16)

We drop the tilde for notational simplicity.

### 2.1.3 Spatial Discretization

For simulation purposes, the PDE-ODE system (2.1) and (2.2), is converted into a system of ODEs. This is achieved by discretizing the PDE-ODE system in space using a standard finite volume method as introduced in [21].

We start by placing an \( N \times M \) uniform grid over the rectangular domain \( \Omega := [0, L] \times [0, W] \). Equation (2.1) is integrated over each grid cell area and using the divergence theorem we have

\[ \frac{d}{dt} \int_{v_{i,j}} M \, dx \, dy = \int_{\partial v_{i,j}} J_n \cdot ds + \int_{v_{i,j}} F(C)M \, dx \, dy, \quad i = 1, \ldots, N, \quad j = 1, \ldots, M, \]  
(2.17)

where the area integral of \( \nabla \cdot (D(M)\nabla M) \) is switched to the surface integral of the outward normal flux, \( J_n = D(M)\partial_n M \), through the boundary of a cell, \( \partial v_{i,j} \). Here \( v_{i,j} \) stands for the area of a grid cell. The biomass density and carbon concentration are both evaluated at the centers of the grid cells, which in terms of the grid indices and grid lengths are defined as

\[ M_{i,j}(t) := M(t, x_i, y_j) \approx M\left(t, \left(i - \frac{1}{2}\right) \Delta x, \left(j - \frac{1}{2}\right) \Delta x\right), \]  
(2.18)

\[ C_{i,j}(t) := C(t, x_i, y_j) \approx C\left(t, \left(i - \frac{1}{2}\right) \Delta x, \left(j - \frac{1}{2}\right) \Delta x\right), \]  
(2.19)

\[^1\text{Here we focus on the two dimensional case.}\]
where the sides of each grid cell are equal to \( \Delta x = \frac{L}{N} = \frac{W}{M} \). The area integrals are evaluated by the mid-point rule. The flux integrals are evaluated for each cell edge separately by the mid-point rule as well. Since the fluxes are computed at the grid cell edges, we need to evaluate the diffusion coefficient at these positions. To this end, we follow [13] where the flux across a shared interface of two grid cells is approximated by the arithmetic average of the diffusion at the centers of those cells. The derivative of \( M \) across each cell edge is approximated by the central finite difference formula. By applying all the above definitions and evaluations to Eq. (2.1) we end up with an ODE for the biomass density in the following form

\[
\frac{d}{dt} M_{i,j} = \frac{1}{\Delta x} \left( J_{i+\frac{1}{2},j} + J_{i-\frac{1}{2},j} + J_{i,j+\frac{1}{2}} + J_{i,j-\frac{1}{2}} \right) + F_{i,j} M_{i,j},
\]

where \( F_{i,j} := F(C_{i,j}) = \frac{C_{i,j}}{\kappa + C_{i,j}} - \lambda \). By imposing the boundary condition (2.6) each flux term in (2.20) is given by

\[
\begin{align*}
J_{i+\frac{1}{2},j} &= \begin{cases} 
\frac{1}{2\Delta x} (D(M_{i+1,j}) + D(M_{i,j})) (M_{i+1,j} - M_{i,j}), & \text{for } i < N, \\
0, & \text{for } i = N,
\end{cases} \\
J_{i-\frac{1}{2},j} &= \begin{cases} 
0, & \text{for } i = 1, \\
\frac{1}{2\Delta x} (D(M_{i,j}) + D(M_{i-1,j})) (M_{i-1,j} - M_{i,j}), & \text{for } i > 1,
\end{cases} \\
J_{i,j+\frac{1}{2}} &= \begin{cases} 
\frac{1}{2\Delta x} (D(M_{i,j+1}) + D(M_{i,j})) (M_{i,j+1} - M_{i,j}), & \text{for } j < M, \\
0, & \text{for } j = M,
\end{cases} \\
J_{i,j-\frac{1}{2}} &= \begin{cases} 
0, & \text{for } j = 1, \\
\frac{1}{2\Delta x} (D(M_{i,j}) + D(M_{i,j-1})) (M_{i,j-1} - M_{i,j}), & \text{for } j > 1.
\end{cases}
\]

Following the same methodology, the spatial discretization is applied to Eq. (2.2). The main difference here is that there are no fluxes across cell edges since we have assumed immobile substrate. Therefore we have

\[
\frac{d}{dt} C_{i,j} = -G_{i,j} C_{i,j},
\]

with \( G_{i,j} := G(C_{i,j}) = \Upsilon \frac{C_{i,j}}{\kappa + C_{i,j}} \).

For computational purposes, the lexicographical grid ordering is introduced as follows

\[
\pi : \{1, \ldots, N\} \times \{1, \ldots, M\} \to \{1, \ldots, NM\}, \quad (i,j) \mapsto p = (i-1)M + j.
\]
Using the above map, we end up with two \( NM \)-component vectors for the biomass density and carbon concentration: \( \mathbf{M} = (M_1, ..., M_{NM})^T \) and \( \mathbf{C} = (C_1, ..., C_{NM})^T \) with \( M_p := M_{\pi(i,j)} = M_{i,j} \) and \( C_p := C_{\pi(i,j)} = C_{i,j} \). Hence, spatial discretization, converts the model PDE-ODE system into a \( 2 \cdot N \cdot M \) system of ODEs, namely

\[
\frac{d\mathbf{M}}{dt} = \mathbf{D}(\mathbf{M}) \mathbf{M} + \mathbf{F}(\mathbf{C}) \mathbf{M},
\]

\[
\frac{d\mathbf{C}}{dt} = -\mathbf{G}(\mathbf{C}) \mathbf{M}.
\]

(2.24)

Here \( \mathbf{D}(\mathbf{M}) \), \( \mathbf{F}(\mathbf{C}) \) and \( \mathbf{G}(\mathbf{C}) \) are \( NM \times NM \) matrices. The entries of \( \mathbf{D}(\mathbf{M}) \) are calculated by the use of the first term in (2.20) and the corresponding definitions in (2.21) after the biomass density indices are mapped to \( p \) using (2.23). The entries of \( \mathbf{F}(\mathbf{C}) \) and \( \mathbf{G}(\mathbf{C}) \) are evaluated by using the growth rate and uptake rate formulae for each \( C_p \), respectively.

**Remark 1.** The matrix \( \mathbf{D}(\mathbf{M}) \) is symmetric and weakly diagonally dominant with non-positive diagonal entries and non-negative off-diagonal entries. The matrices \( \mathbf{F}(\mathbf{C}) \) and \( \mathbf{G}(\mathbf{C}) \) are diagonal. For \( C_p = 0 \) the \( p \)th diagonal entry of \( \mathbf{G}(\mathbf{C}) \) becomes zero while the same entry in \( \mathbf{F}(\mathbf{C}) \) becomes negative with a magnitude equal to the cell loss rate.

### 2.1.4 Regularization of the Model Equations

The standard arguments in the theory of ordinary differential equations that may be used to prove the existence, uniqueness and boundedness of solutions in a specific interval for physical reasons, are not applicable to the system (2.24). This is mainly due to the existing singularity in the density dependent diffusion defined by (2.3). One way around this problem is to regularize the PDE Eq. (2.1) and hence the corresponding semi-discretized ODE system (2.24). For this purpose, we introduce the regularized version of the diffusion coefficient (2.3) [16]

\[
D_\epsilon(M) = \begin{cases} 
\delta(M + \epsilon)^{\alpha}, & M \leq 1 - \epsilon, \\
\delta\epsilon^{-\beta}, & M \geq 1 - \epsilon,
\end{cases}
\]

(2.25)

with regularization factor \( 0 < \epsilon \ll 1 \). By assuming the regularized diffusion in this form, the regularized PDE-ODE equations (2.1) and (2.2) and the corresponding semi-discretized
system of ODEs (2.24) can be written as
\[
\begin{aligned}
\partial_t M_\epsilon &= \nabla \cdot (D_\epsilon (M_\epsilon) \nabla M_\epsilon) + F(C_\epsilon) M_\epsilon, \\
\partial_t C_\epsilon &= -G(C_\epsilon) M_\epsilon,
\end{aligned}
\]  
(2.26)

and
\[
\begin{aligned}
\frac{dM^\epsilon}{dt} &= \mathcal{D}_\epsilon (M^\epsilon) M^\epsilon + \mathcal{F} (C^\epsilon) M^\epsilon, \\
\frac{dC^\epsilon}{dt} &= -\mathcal{G} (C^\epsilon) M^\epsilon.
\end{aligned}
\]  
(2.27)

respectively. Here, \(\epsilon\) is used as subscripts and superscripts to signify that the biomass density and the carbon concentration in these equations belong to the regularized problem.

It has been shown that the solutions of (2.26) with diffusing substrate converge to the solutions of the original PDE-ODE system (2.1) and (2.2) as \(\epsilon \to 0\). Also, the biomass density of the regularized problem is bounded from above i.e. \(M_\epsilon \leq 1 - \eta\) for some \(\eta > 0\) [16]. The existence, uniqueness, non-negativity and boundedness of the solutions of the regularized problem (2.27) with diffusing substrate, initial data in \([0, 1]\) and sufficiently small \(\epsilon\) has been proven in [21]. The convergence of the solutions as \(\epsilon \to 0\) to the solutions of (2.24) is given as a theorem and its proof in [21] as well. We assume the existence, uniqueness, non-negativity and boundedness of the solutions of the regularized problems (2.26) and (2.27) will follow from similar arguments provided in [16] and [21] adapted to the case of non-diffusing substrate, and that their solutions converge to those of the original PDE-ODE system as \(\epsilon \to 0\).

The description of the deterministic spatially explicit continuum model proposed in [12] is now complete. In the next section we introduce our approach to including stochastic attachment of cells within the framework of the deterministic model detailed in this section.

2.2 A Spatially Explicit Model with Random Cell Attachments

As mentioned in the introductory chapter, we have developed a new approach, using the theory of stochastic differential equations, to model the stochastic attachment of cells to the substratum from the aqueous phase. We discard model assumption 6 in Section 2.1.1 to allow the occurrence of attachments. We start the mathematical description of our approach
by reviewing stochastic differential equations, especially the formalism of Itô SDEs. This is followed by the addition of an impulse function to the governing equations of cellulolytic biofilms. We show how the proposed Itô SDE is coupled to the modified semi-discrete system of ODEs to make a single system of SDEs and finally we examine the non-negativity and boundedness of this system in its regularized format.

2.2.1 Stochastic Differential Equations

Here we focus on stochastic differential equations (SDEs) since this is the formalism that we will use in our approach later.

We are familiar with the mathematical notation of an initial value problem (IVP), namely

\[
\frac{dX(t)}{dt} = f(t, X(t)), \quad t \in (t_0, \infty),
\]

\[
X(t_0) = X_0, \tag{2.28}
\]

where \( f : [0, \infty) \times \mathbb{R}^n \mapsto \mathbb{R}^n \). The stochastic version of the above problem is written as \([47, 29]\)

\[
\frac{dX(t)}{dt} = f(t, X(t)) + g(t, X(t)) \cdot \xi_t, \quad t \in (t_0, \infty),
\]

\[
X(t_0) = X_0, \tag{2.29}
\]

where the original IVP is perturbed by a noise term composed of \( g : [0, \infty) \times \mathbb{R}^n \mapsto \mathbb{R}^n \) and \( \xi_t \), which represents standard Gaussian random variables for each \( t \). It turns out the stochastic process that can represent \( \xi_t dt \) is the Gaussian white noise process: a stochastic process of normally distributed random variables with zero mean, zero correlation at different times and finite variance \([47, 29]\). We can rewrite the first line of (2.29) in the following form

\[
dX(t) = f(t, X(t))dt + g(t, X(t)) \cdot \xi_t dt. \tag{2.30}
\]

The stochastic process that satisfies the properties of white noise and whose increments can replace \( \xi_t dt \) in (2.30), is Brownian motion. In the mathematical framework Brownian motion is represented by the standard Wiener process \( W_t \) (see Figure 2.1), which is a Gaussian process with the following properties \([29]\)

1. \( W_0 = 0 \) almost surely (with probability 1).

2. Independent increments \( e.g. W_{t_1} - W_{s_1} \) and \( W_{t_2} - W_{s_2} \) are independent random variables
for all $0 \leq s_1 < t_1 \leq s_2 < t_2$.

3. The increments are normally distributed with $E(W_t - W_s) = 0$ and $Var(W_t - W_s) = t - s$
   for all $0 \leq s < t$

4. Its sample paths are continuous functions of time with probability 1.

Figure 2.1: A representation of the sample path of a standard Wiener process, $W_t$, with properties 1-4.

A formal definition of a system of Itô SDEs where the stochasticity is driven by standard Wiener processes is given as follows: consider a probability space such as $(\Omega, \mathcal{A}, P)$ consisting of the sample space $\Omega$, the collection of events $\mathcal{A}$ also known as a $\sigma$-algebra and the probability measure $P$. Furthermore, the probability space contains an increasing right-continuous family of $\sigma$-algebras $\mathcal{A} = \{\mathcal{A}_t, t \geq 0\}$, which is interpreted as information known until time
The system of Itô SDEs is formulated as
\[
dX(t) = f(t, X(t))dt + g(t, X(t))dW_t, \quad t \in (t_0, \infty),
\]
\[
X(t_0) = X_0.
\]
where \( f : [0, \infty) \times \mathbb{R}^n \mapsto \mathbb{R}^n \) is a function called the drift coefficient, \( g : [0, \infty) \times \mathbb{R}^n \mapsto \mathbb{R}^{n \times m} \) is a matrix that is known as the diffusion coefficient (in the stochastic context) and \( dW_t \) is the Itô differential, which is an infinitesimal increment of \( W : [0, \infty) \times \Omega \mapsto \mathbb{R}^m \), an \( m \)-dimensional, \( A \)-adapted (its values are detectable by events in \( A \)) Wiener process. Here, the initial time, \( t_0 \), is non-negative and the initial data is \( X_0 \in \mathbb{R}^n \) [9, 29]. The Wiener process is nowhere differentiable and of unbounded variation. Consequently Eq. (2.31) is recognized as the formal notation of Itô SDEs in the literature. The rigorous form of the stochastic IVP (2.31) is formulated by the integral equation
\[
X(t) = X_0 + \int_{t_0}^{t} f(s, X(s))ds + \int_{t_0}^{t} g(s, X(s))dW_s, \quad t \in [t_0, \infty).
\]
Here, the first term is an ordinary (Riemann or Lebesgue) integral and the second term is an Itô integral, which requires the Itô calculus formalism for evaluation [47, 29].

Two well-known types of noises are additive noise and multiplicative noise: in the case of additive noise, \( g(t, X(t)) \) is a real constant and the noise is called multiplicative if \( g(t, X(t)) = aX(t) \) for some real constant \( a \).

It must be mentioned that besides the Itô’s definition of stochastic integrals, we have Stratonovich’s definition (see Chapter 3 in [47] for this definition) as another alternative for rigorously defining (2.32). For our modeling purposes we have used Itô’s definition and we assume that using Stratonovich’s definition should be straightforward as well but we have not applied it here.

## 2.2.2 Mathematical Formulation of Stochastic Cell Attachment

The outline of our approach to the modeling of the stochastic attachment of bacteria is provided in Section 1.6. Here we put our approach in mathematical terms.

We start by adding an impulse function to the biomass ODE (2.20) in order to devise a mechanism for picking up the stochastic attachment signals received as inputs and converting these signals into attached biomass at some arbitrary position on the substratum. These input signals, which we call the attachment factors (AFs), are denoted by \( \phi \) in the upcoming
equations. There are different options for the impulse function, \( f(\phi) \). Our goal is to choose the best candidate for this function (a) with properties that allow us to control different aspects of the attachment and (b) that has suitable characteristics for numerical simulations. For instance the normal distribution equation possesses most of the required properties however its numerical realization within this modeling framework is not well-behaved \( e.g. \) its peak value is dependent on its width for which some issues with the upper bound of the biomass density may arise. This is obviously not desired since we need to have control over the attached biomass density. For this reason alone, choosing an impulse function with a maximum equal to the theoretical upper bound of the biomass density could be considered a good choice. We have constructed an impulse function from the Fermi-Dirac distribution mostly used in physics, which lends itself very well to our pursued modeling framework. The range of this function is \([0, 1]\) and its shape can be adjusted by tweaking its characterizing parameters \( a, b \) and \( \sigma \)

\[
f(\phi) = \frac{1}{e^{\frac{a}{\sigma} - b} + 1} - \frac{1}{e^{\frac{a}{\sigma} + b} + 1}, \quad a \in \mathbb{R}, \quad b \in \mathbb{R}, \quad \sigma > 0.
\] 

(2.33)

In Figure 2.2 several instances of \( f(\phi) \) are plotted for \( a = 0.9 \) and \( b = 1.0 \). One can readily notice that as the value of \( \sigma \) is reduced, the function becomes more rectangular and its support approaches \([0.9, 1.0]\) from left and right.

The biomass ODE (2.20) is rewritten using the introduced impulse function as the following

\[
\frac{dM_{i,j}}{dt} = \left( J_{i+\frac{1}{2},j} + J_{i-\frac{1}{2},j} + J_{i,j+\frac{1}{2}} + J_{i,j-\frac{1}{2}} \right) + F_{i,j}M_{i,j} + \gamma_1(1 - M_{i,j})^{\gamma_2}C_{i,j}^{\gamma_3}f_{i,j},
\]

\[
0 < \gamma_1 \ll 1, \quad \gamma_2 \geq 1, \quad \gamma_3 \geq 1,
\]

(2.34)

where all the terms except the last one follow the definitions given for (2.20) and (2.21), and \( f_{i,j} := f(\phi_{i,j}) = (e^{\frac{\phi_{i,j}-b}{\sigma}} + 1)^{-1} - (e^{\frac{\phi_{i,j}+a}{\sigma}} + 1)^{-1}. \)

Here we have introduced the coefficient, \( \gamma_1(1 - M_{i,j})^{\gamma_2}C_{i,j}^{\gamma_3} \). The factor, \( (1 - M_{i,j})^{\gamma_2} \) accounts for the existing biomass so that the biomass density stays within its bounds and \( C_{i,j}^{\gamma_3} \) prevents attachment to the carbon-free positions. The exponents \( \gamma_2 \) and \( \gamma_3 \) are included to introduce a generic formulation of the coefficient. Moreover, \( \gamma_1 \) provides the means to control the density of bacterial attachment per unit time. It should be noted that its value is chosen to be much smaller than \( 1 \) so that the numerical realizations would be in agreement with the experimental observations in [11] and [57].
Figure 2.2: The impulse function, $f(\phi)$, plotted for different values of the parameter $\sigma$ where $a = 0.9$ and $b = 1.0$.

The spatially discretized version of the substrate ODE (2.22) stays the same as before.

In order to simulate the AF for each grid cell, the Itô SDE formalism detailed in Section 2.2.1 is deployed. The governing equation reads

$$d\phi_{i,j} = \psi(1 - M_{i,j})^{\nu_1} C_{i,j}^{\nu_2} dW_t^{(i,j)}, \quad \nu_1 \geq 1, \quad \nu_2 \geq 1,$$

(2.35)

where the drift term is zero\(^2\) ($0 \cdot dt$) and $dW_t^{(i,j)}$ is the Itô differential for the cell designated by $(i,j)$. Here, exponents $\nu_1$ and $\nu_2$ are included for the sake of generalization and the constant coefficient, $\psi$, controls the magnitude of noise, which can be harnessed as a control parameter for the rate of attachment events. The diffusion coefficient, $\psi(1 - M_{i,j})^{\nu_1} C_{i,j}^{\nu_2}$, causes the stochastic fluctuations in the system to be proportional to volume filling and available substrate. Its incorporation in the present form is a bit intuitive and one could justify it by arguing that the amplitude of stochastic variation that is manifested by $\phi_{i,j}$

\(^2\)Since we have no justification for assuming otherwise.
can be linked to the available space and substrate in a grid cell so that the magnitude of fluctuations in this variable are effected by the other state variables in a causal manner. To clarify this proposition and make it more rigorous, we make the following arguments: the attachment events occur when the value of \( \phi_{i,j} \) in a grid cell falls within the support of the impulse function (assuming that initially \( \phi_{i,j} \) is not within the support of \( f_{i,j} \)); the more frequently this happens, the higher could be the rate of attachment; obviously when a grid cell is free of biomass and contains the highest concentration of carbon, the greater is the amplitude of stochastic variation in (2.35), which increases the likelihood for \( \phi_{i,j} \) to acquire a value within the support of \( f(\phi_{i,j}) \), which in turn increases the chance of attachment of bacteria to that grid cell. As time progresses, either the biomass density reaches its saturation level or the carbon concentration tends to zero. In either case, the right hand side of (2.35) becomes zero and \( \phi_{i,j} \) assumes a steady state\(^3\).

Recognizing the mechanism whereby the assumed form for (2.35) operates, one could argue that even if the stochastic variation tends to zero for a cell as space is occupied and substrate is consumed, there is a non-zero probability that the asymptotic (steady state) value of \( \phi_{i,j} \) falls and stays within the support of \( f_{i,j} \). Although this is true and inevitable due to the particular formulation of (2.35), the coefficient, \( \gamma_1 (1 - M_{i,j})^{\gamma_2} C_{i,j}^{\gamma_3} \), of the impulse function nullifies the lasting impulse that can be caused by these relatively rare events: its dependence on carbon concentration switches off the attachment of any bacteria when the supply of carbon in a grid cell is depleted to zero.

The semi-discrete system of equations for this particular approach is obtained via the lexicographical grid ordering (2.23) in the same way we used it to obtain (2.24). The major difference is the addition of the impulse function to the biomass ODEs and the introduction of the Itô SDEs (2.35). We make the \( i, j \)-designation for the AF and the Itô differential precisely defined by indicating that these quantities are both evaluated at the centers of the grid cells, which in terms of the grid indices and grid lengths are defined as

\[
\phi_{i,j}(t) := \phi(t, x_i, y_j) \approx \phi \left( t, \left( i - \frac{1}{2} \right) \Delta x, \left( j - \frac{1}{2} \right) \Delta x \right), 
\]

\[
dW_{t}^{(i,j)} := dW(t, x_i, y_j) \approx dW \left( t, \left( i - \frac{1}{2} \right) \Delta x, \left( j - \frac{1}{2} \right) \Delta x \right).
\]

Using the mapping (2.23), we have two more \( NM \)-component vectors for the AF and the \(^3\)When the biomass density reaches its maximum value the corresponding steady state is going to be relatively short-lived since the diffusion and cell loss cause the biomass value to fall below the saturation level at some point.
Itô differential in addition to the biomass density and the substrate concentration vectors: \( \phi = (\phi_1, ..., \phi_{NM})^T \) and \( dW_t = \left( dW_t^{(1)}, ..., dW_t^{(NM)} \right)^T \) with \( \phi_p := \phi_{\pi(i,j)} = \phi_{i,j} \) and \( dW_t^{(p)} := dW_t^{(\pi(i,j))} = dW_t^{(i,j)} \).

The whole system comprises \( 3 \cdot N \cdot M \) SDEs where the biomass and carbon ODEs are the special cases of Itô SDEs for which the stochastic terms are identically zero\(^4\). The system reads

\[
\begin{align*}
\frac{dM}{dt} &= \left[ D(M) M + \mathcal{I}(C) M + \mathcal{J}(M, C) F(\phi) \right] dt, \\
\frac{dC}{dt} &= \left[ -\mathcal{G}(C) M \right] dt, \\
\frac{d\phi}{dt} &= \mathcal{T}(M, C) dW_t,
\end{align*}
\]

(2.38)

where \( \mathcal{D}(M) \), \( \mathcal{I}(C) \) and \( \mathcal{G}(C) \) are \( NM \times NM \) matrices with properties described in Remark 1. We have introduced two new \( NM \times NM \) matrices: \( \mathcal{I}(M, C) \) and \( \mathcal{T}(M, C) \). The vector \( F(\phi) \) has \( NM \) components and its entries are the impulse function evaluated for each grid cell.

**Remark 2.** The matrices \( \mathcal{I}(M, C) \) and \( \mathcal{T}(M, C) \) are diagonal and their \( p \)th diagonal entries are given by \( \gamma_1 (1 - M_p) \gamma_2 C_p^{\gamma_3} \) and \( \psi(1 - M_p) \gamma_1 C_p^{\gamma_2} \), respectively. For either \( M_p = 1 \) or \( C_p = 0 \), the corresponding diagonal element vanishes. The maximum values on the diagonal are attained when \( M_p = 0 \) and \( C_p = 1 \). Therefore, we have \( 0 \leq i_{pp} \ll 1 \) and \( t_{pp} \geq 0 \) where \( i_{pp} \) and \( t_{pp} \) are the \( p \)th diagonal entries of \( \mathcal{I}(M, C) \) and \( \mathcal{T}(M, C) \), respectively. The components of the vector \( F(\phi) \) are given by \( f_p := f_{\pi(i,j)} = f_{i,j} \) such that \( f_p = (e^{\phi_p \frac{\delta}{\sigma}} + 1)^{-1} - (e^{\phi_p \frac{-\delta}{\sigma}} + 1)^{-1} \).

By construction, we have \( 0 \leq f_p \leq 1 \) where the \( p \)th entry vanishes when the value of the \( p \)th entry of \( \phi \) is not within the support of the impulse function.

We recall from Section 2.1.4 that due to the singularity in the diffusion coefficient (2.3), we cannot determine the general behavior of the solutions (assuming their existence and uniqueness) of (2.38), such as non-negativity and boundedness. In order to prepare the system for investigations of this type, we utilize a regularized diffusion coefficient similar to (2.25) that reads

\[
D_\epsilon(M) = \begin{cases} 
0, & M < 0, \\
\delta \frac{M^\alpha}{(1 - M)^\beta}, & 0 \leq M \leq 1 - \epsilon, \\
\delta \epsilon^{-\beta} (1 - \epsilon)^\alpha, & M > 1 - \epsilon,
\end{cases}
\]

(2.39)

\(^4\) An SDE with no noise term is a regular ODE, therefore for the sake of the simplification of terminology, we describe the whole system of equations as a collection of SDEs.
where we have extended $D_\epsilon(M)$ for negative values of $M$. Consequently, the regularized version of the SDE problem (2.38) is

$$dM^\epsilon = \left[ D_\epsilon(M^\epsilon) M^\epsilon + \mathcal{F}(C^\epsilon) + I(M^\epsilon, C^\epsilon) F(\phi^\epsilon) \right] dt,$$

$$dC^\epsilon = \mathcal{G}(C^\epsilon) M^\epsilon dt,$$

$$d\phi^\epsilon = \mathcal{T}(M^\epsilon, C^\epsilon) dW_t. \quad (2.40)$$

### 2.2.3 Non-negativity and Boundedness in the Regularized Problem

Assuming the existence and uniqueness of the solutions for the biomass density and the carbon concentration in the regularized system of SDEs (2.40), we want to show whether these solutions are positive and bounded from above. To this end, we utilize the invariant set theorem for systems of SDEs (detailed in [9]), which in the limit of zero stochastic perturbations is equivalent to the tangent condition for systems of ODEs (see [56]). We start by defining an invariant set in the context of both ODEs and SDEs. Then we recall a theorem from [9] that provides the necessary and sufficient conditions for a system of SDEs to maintain an invariant set containing the solutions. We use the ODE and SDE systems (2.28) and (2.31) as references in the following statements.

**Definition 1.** $S \subset \mathbb{R}^n$ is called an invariant set for the stochastic system (2.31) if for every initial data $X_0 \in S$ and initial time $t_0 \geq 0$, the corresponding solution $X(t)$ with $t \geq t_0$ almost surely remains in $S$, or equivalently

$$P \left( \{ X(t) \in S, t \in [t_0, \infty) \} \right) = 1.$$

In the case the coefficient of the stochastic term in (2.31) is set to zero, the SDE system becomes an ODE system for which the definition of an invariant set is given below.

**Definition 2.** $S \subset \mathbb{R}^n$ is called an invariant set for the ODE system (2.28) if for every initial data $X_0 \in S$ and initial time $t_0 \geq 0$, the corresponding solution $X(t)$ with $t \geq t_0$ remains in $S$, or equivalently

$$\forall X_0 \in S \Rightarrow X(t) \in S, t \in [t_0, \infty)$$
Theorem 1. Let $a_p, b_p \in \mathbb{R} \cup \{\infty\}, b_p > a_p$ and $p \in I$ where $I \subseteq \{1, \ldots, n\}$ is non-empty. The set

$$S := \{X \in \mathbb{R}^n : a_p \leq X_p \leq b_p, p \in I\}$$

is invariant for the SDE system (2.31) if and only if

- $f_p(t, X) \geq 0$ for $X \in S$ such that $X_p = a_p$,
- $f_p(t, X) \leq 0$ for $X \in S$ such that $X_p = b_p$,
- $g_{p,r}(t, X) = 0$ for $X \in S$ such that $X_p \in \{a_p, b_p\}, r = 1, \ldots, m,$

for all $t \geq 0$ and $p \in I$.

This theorem is equivalent to the tangent condition for systems of ODEs when the coefficient of the stochastic term i.e. $g$, is identically zero. Therefore, in the particular case of $g = 0$, the set $S$ is invariant for the ODE system (2.28) if and only if the conditions involving $f$ are satisfied.

The next step is to show whether the biomass density and the carbon concentration solutions remain within their physically valid range of values. Mathematically, we want to show there exist certain invariant sets, as defined above, for $M^\epsilon$ and $C^\epsilon$. In the following statements we make use of vector $1 \in \mathbb{R}^{NM}$ and $0 \in \mathbb{R}^{NM}$ where $1 = (1, \ldots, 1)^T$ and $0 = (0, \ldots, 0)^T$. Furthermore, with an inequality for vectors such as $U \leq V$ we mean $U_p \leq V_p$ for $\forall p \in \{1, \ldots, NM\}$.

Proposition 1. Let $M^\epsilon$ and $C^\epsilon$ be solution components of (2.40). Suppose the initial data for this system of SDEs satisfy $0 \leq M^\epsilon(0) \leq \eta 1$ with $\eta \in (0, 1)$ and $0 \leq C^\epsilon(0) \leq 1$. Then for sufficiently small $\epsilon > 0$, the solutions are non-negative and bounded by their physical limit i.e. $0 \leq M^\epsilon \leq 1$ and $0 \leq C^\epsilon \leq 1$.

Proof. The biomass density and the carbon concentration are governed by the first and second equations in (2.40). These equations are instances of SDE systems where the stochastic term, $g$, is identically zero. Therefore, the application of Theorem 1 is equivalent to checking the tangent condition for ODEs. We start by examining the carbon concentration equation. Given the initial data, $0 \leq C^\epsilon(0) \leq 1$, we evaluate the right hand side of the second equation in (2.40) at the lower and upper bounds of $C^\epsilon$. According to Remark 1, the matrix $\mathfrak{G}(C^\epsilon)$ is diagonal. Thus, the $p$th component of $\frac{d}{dt}C^\epsilon$ is given by the $p$th component of $-\mathfrak{G}(C^\epsilon)M^\epsilon$ i.e. $-\Upsilon_{\kappa + C^\epsilon} C^\epsilon_{kp} M^\epsilon_{kp}$. For $C^\epsilon_p = 0$ the latter vanishes and for $C^\epsilon_p = 1$ it is non-positive if $M^\epsilon_p \geq 0$. Therefore, by Theorem 1, $0 \leq C^\epsilon$ is guaranteed.
The biomass equation in (2.40) can be divided into three parts: the first part involves the contribution from the stochastic cell attachment, \( J(M^e, C^e) F(\phi^e) \); the second part accounts for the biomass net growth, \( \mathcal{F}(C^e) M^e \); the third part quantifies the biomass diffusion \( \mathcal{D}_e(M^e) M^e \). As a result, the \( p \)th component of \( \frac{d}{dt} M^e \) is the sum of these parts. Recalling Eq. (2.21), Remark 1 and Remark 2, we have

- **\( p \)th component of** \( J(M^e, C^e) F(\phi^e) \):
  \[
  i_{f_p} := \gamma_1 (1 - M^e_p)^{\gamma_2} C^e_p f_p,
  \]

- **\( p \)th component of** \( \mathcal{F}(C^e) M^e \):
  \[
  f_{m_p} := \left( \frac{C^e_p}{\kappa + C^e_p} - \lambda \right) M^e_p,
  \]

- **\( p \)th component of** \( \mathcal{D}_e(M^e) M^e \):
  \[
  \varnothing_{m_p} := \frac{1}{2\Delta x} \left[ \left( D_e(M^e_{p-M}) + D_e(M^e_p) \right) M^e_{p-M} + \left( D_e(M^e_{p-1}) + D_e(M^e_p) \right) M^e_{p-1} \right. \\
  - \left( D_e(M^e_{p-M}) + D_e(M^e_{p-1}) + 4D_e(M^e_p) + D_e(M^e_{p+1}) + D_e(M^e_{p+M}) \right) M^e_p \\
  + \left( D_e(M^e_{p+1}) + D_e(M^e_p) \right) M^e_{p+1} + \left( D_e(M^e_{p+M}) + D_e(M^e_p) \right) M^e_{p+M} \right].
  \]

For \( M^e_p = 0 \), we have \( 0 \leq i_{f_p} \leq K \gamma_1 (K := \max(C^e \gamma_3, 1)) \) and \( f_{m_p} = 0 \). The diffusive part is given by

\[
\varnothing_{m_p} = \frac{1}{2\Delta x} \left[ \left( D_e(M^e_{p-M}) \right) M^e_{p-M} + \left( D_e(M^e_{p-1}) \right) M^e_{p-1} \right. \\
+ \left( D_e(M^e_{p+1}) \right) M^e_{p+1} + \left( D_e(M^e_{p+M}) \right) M^e_{p+M} \right].
\]

If the four neighbors of \( M^e_p \), namely \( M^e_{p-M}, M^e_{p-1}, M^e_{p+1} \) and \( M^e_{p+M} \), are non-negative, then we have \( \varnothing_{m_p} \geq 0 \). Given the initial data, \( 0 \leq M^e(0) \leq \eta \mathbf{1} \), for any neighbor to become negative, it has to continuously go through zero. Since there is zero net growth as well as non-negative flux into a neighboring cell with zero biomass, it is impossible for the neighbors of \( M^e_p \) to attain negative values. Hence, \( \varnothing_{m_p} + f_{m_p} + i_{f_p} \geq 0 \) leading to \( M^e \) being non-negative, which also guarantees \( C^e \leq 1 \).
For \( M^\epsilon_p = 1 \), we have \( \text{if}_p = 0 \) and \( f_m_p \leq Q \) with \( Q := \frac{1}{1+\kappa} \). The diffusive part is given by

\[
\begin{align*}
\partial m_p &= \frac{1}{2\Delta x} \left[ \left( D(\epsilon(M^\epsilon_{p-M}) + \delta \epsilon^{-\beta} (1-\epsilon) \alpha) M^\epsilon_{p-M} + (D(\epsilon(M^\epsilon_{p-1}) + \delta \epsilon^{-\beta} (1-\epsilon) \alpha) M^\epsilon_{p-1}
\right.

- \left. \left( D(\epsilon(M^\epsilon_{p-M}) + \delta \epsilon^{-\beta} (1-\epsilon) \alpha) M^\epsilon_{p-M} + 4\delta \epsilon^{-\beta} (1-\epsilon) \alpha + D(\epsilon(M^\epsilon_{p+1}) + D(\epsilon(M^\epsilon_{p+M}))
\right.

+ \left. \left( D(\epsilon(M^\epsilon_{p+1}) + \delta \epsilon^{-\beta} (1-\epsilon) \alpha) M^\epsilon_{p+1} + (D(\epsilon(M^\epsilon_{p+M}) + \delta \epsilon^{-\beta} (1-\epsilon) \alpha) M^\epsilon_{p+M} \right) \right].
\end{align*}
\]

According to Remark 1, matrix \( \mathcal{D}(M^\epsilon) \) is weakly diagonally dominant. Therefore, the negative part of the above equation is either equal or greater than the positive part in magnitude. An estimate for \( \partial m_p \) can be written as

\[
-\frac{2\delta \epsilon^{-\beta} (1-\epsilon) \alpha}{\Delta x} \leq \partial m_p \leq \frac{\delta \epsilon^{-\beta} (1-\epsilon) \alpha (\Sigma - 4)}{\Delta x},
\]

where \( \Sigma = M^\epsilon_{p-M} + M^\epsilon_{p-1} + M^\epsilon_{p+1} + M^\epsilon_{p+M} \) and initially \( \Sigma < 4 \). If we choose \( \epsilon \) sufficiently small such that \( |\partial m_p| \geq Q \) everywhere, it is guaranteed that \( \partial m_p + f_m_p + \text{if}_p \leq 0 \) at all times. Thus, there is an upper bound for the biomass density i.e. \( M^\epsilon \leq 1 \).

Therefore, according to Theorem 1, \([0,1]\) satisfies the necessary and sufficient conditions to be an invariant set for this SDE system and we have \( 0 \leq C^\epsilon \leq 1 \) and \( 0 \leq M^\epsilon \leq 1 \). \( \square \)
Chapter 3

Computational Implementation

This chapter provides the details of the computational methods we deployed in order to simulate the system of SDEs (2.38).

Following the description of the numerical methods, we provide the computer algorithm for the simulations and we show how the algorithm can be adapted to parallel computing. Furthermore, we discuss in detail the impact of parallelization on simulation speed and analyze the results.

The random number generation for simulating the Wiener process is performed by known algorithms, which we address briefly. We also explain how we parallelize these algorithms in order to obtain more efficiency.

One of the challenges for simulating the explicit numerical methods used here is choosing sufficiently small time steps so that the numerical approximations remain stable and converge. We summarize some of the methods deployed in our computer codes to address these challenges.

3.1 Numerical Methods

We have used the Euler and Euler-Maruyama methods to numerically integrate (2.38). The Euler method is a first-order explicit numerical integration recipe for solving ODEs with given initial values. It is explicit since it calculates the current value of the dependent variable from its past value and it is first-order because its global truncation error scales with the first power of the time step size \( i.e. \ error \sim O(h) \) where \( h \) is the length of the time step [18].
The scheme for the Euler method is commonly written as

\[ Y_{n+1} = Y_n + hf(t_n, Y_n), \]  

(3.1)

where the time discretization is such that \( t_0 < t_1 < t_2 < ... < t_n < t_{n+1} < ... \) and the time increment is \( h = t_{n+1} - t_n \). The time at the \( nth \) step can be calculated using the step size and the initial time: \( t_n = t_0 + nh \). By these definitions, \( Y_{n+1} \) is the approximation to the solution at \( t = t_{n+1} \) and \( Y_n \) is the approximation at \( t = t_n \). Starting from an initial data, this formula is applied recursively to an IVP, such as the one given by (2.28), to find an approximate solution numerically.

The stochastic equivalent of the Euler method, the Euler-Maruyama method, resembles the numerical approximation (3.1) but it also accounts for the contribution from the noise term. This method can be used to numerically calculate approximate sample solutions to Itô SDEs such as (2.31). It has been shown that the Euler-Maruyama method approximates the sample solutions of an SDE with the strong order of convergence equal to 0.5 [30]. The numerical scheme for this method reads

\[ Y_{n+1} = Y_n + hf(t_n, Y_n) + \delta W_n g(t_n, Y_n). \]  

(3.2)

Here, \( \delta W_n \) represents the increment of the Wiener process at the \( nth \) time step. According to the properties of the Wiener process, the Wiener increments are independent Gaussian random variables with the expected value \( E(\delta W_n) = 0 \) and the standard deviation \( \text{Var}(\delta W_n) = h \). These characteristics of the increments make them suitable for being simulated by standard pseudo-random number generators (PRNGs). It should be noted that if the diffusion term, \( g \), is identically zero, the Euler-Maruyama scheme transforms into the usual Euler scheme (3.1).

The computer simulations of (2.38) require the implementation of the schemes (3.1) and (3.2) for which we need to evaluate the corresponding \( f \) and \( g \) in the equations of the biomass density, the carbon concentration and the attachment factor at each time step. This part of the numerical calculations is computationally expensive since it has to be executed for each grid cell separately. In terms of computer code, it means we have to loop through all the grid cells and compute the different quantities that characterize the biomass growth and transport, the carbon consumption and the stochastic attachment of cells per each iteration. The algorithm for performing these computations has the general form as follows:
**input:** The biomass density $M$, carbon concentration $C$ and attachment factor $\phi$.

Each of these is stored in two separate arrays with size $N \times M$: one for the previous time step; one for the current time step

1. Initialize the current and previous arrays for $M$, $C$ and $\phi$;
2. $t \leftarrow t_i$, $T \leftarrow t_f$, $h \leftarrow$ length of time step;
3. // $t_i$ and $t_f$ are the simulation start and stop times.
4. while $t \leq T$ do
   5.     for $i \leftarrow 1$ to $N$ do // loop through $N \times M$ grid cells
      6.         for $j \leftarrow 1$ to $M$ do
      7.             $p \leftarrow (i-1)N + j$;
      8.                 // Implement the lexicographical grid ordering by calculating $p$.
      9.                 if the $p$th cell is on the boundary then
         10.                     determine which cell edges have zero biomass flux;
         11.                 end
      12.             biomass flux $\leftarrow$ left flux + right flux + top flux + bottom flux;
      13.             // biomass flux is calculated using (2.21) for the $p$th grid cell.
      14.             net growth $\leftarrow \left( \frac{C_\text{prev}}{\kappa^\text{prev} + C_p^\text{prev}} - \lambda \right) M_p^\text{prev}$;
      15.             uptake $\leftarrow - \left( \Upsilon \frac{C_\text{prev}}{\kappa^\text{prev} + C_p^\text{prev}} \right) M_p^\text{prev}$;
      16.             cell attachment $\leftarrow \gamma_1(1 - M_p^\text{prev})^\gamma_2 C_p^\text{prev} \gamma_3 f_p$;
      17.             $\delta W_p \leftarrow$ PRNG;
      18.             $M_p^\text{curr} \leftarrow M_p^\text{prev} + (\text{biomass flux} + \text{net growth} + \text{cell attachment}) \times h$;
      19.             $C_p^\text{curr} \leftarrow C_p^\text{prev} + \text{uptake} \times h$;
      20.             $\phi_p^\text{curr} \leftarrow \phi_p^\text{prev} + \psi(1 - M_p^\text{prev})^\mu C_p^\text{prev} \times \delta W_p$;
   5.         end
   21.     end
   22.     $t \leftarrow t + h$;
   23.     copy the values stored in current arrays into previous arrays
24. end

**Algorithm 1:** This is the algorithm for the computer simulation of the system of SDEs (2.38).

It should be noted that the above algorithm does not include all the details of the actual computer code we used for the simulations. Depending on the programming language, different approaches for translating Algorithm 1 into computer code may be adopted. Our simulations were written in the C language and compiled by the Intel compiler. As a result we adapted Algorithm 1 to the syntax and semantics of C. For instance, certain capabilities of the C language, such as the dynamic memory allocation and the pointer operations, make
the memory address swap the best approach to handle the last line of Algorithm 1.

### 3.2 Parallelization

The structure of Algorithm 1 makes it a good candidate for parallel computing and hence boosting the run-time speed of the simulations. The basic idea for parallelization here is to divide the loop iterations between the physical threads of the central processing units (CPUs) provided by the computational hardware. We chose OpenMP, an application programming interface (API), to parallelize our computer code. The primary reason for our choice is the simplicity of the OpenMP framework in comparison with other APIs such as Message Passing Interface (MPI). Moreover, we deemed OpenMP most suitable for our purposes since it can be implemented by applying a few modifications to the serial version of the source code [25]. Overall, the parallelization of computer programs by OpenMP is achieved via a set of compiler directives, runtime library routines and environment variables [48].

We have utilized the work-sharing constructs\(^1\) within the OpenMP framework to execute the outer loop in Algorithm 1 (line 5) in parallel. Using compiler directives such as `#pragma omp for`, each thread inside the parallel region is assigned a “chunk” of loop iterations.

We measured the time the while-loop (Algorithm 1 line 4) takes to complete using OpenMP library routine `omp_get_wtime()`. We did not concern ourselves with the time for initialization and I/O processes, which can be parallelized as well, since the bulk of computational work takes place inside the while-loop. In order to quantify the amount of gained speedup due to parallelization, we used the following simple formula

\[
speedup = \frac{1}{\frac{t_n}{t_s}}, \tag{3.3}
\]

where \(t_n\) is the elapsed time for \(n\) threads to execute the task and \(t_s\) is the elapsed time in the serial (single thread) configuration to execute the same task. The ratio \(\frac{t_n}{t_s}\) is therefore the fraction of time taken by \(n\) threads to execute the same block of code.

In order to have a better understanding of the speedup results, we have provided the hardware and software specifications below:

---

\(^1\)Operations where independent work is assigned to a number of threads for processing [48].
<table>
<thead>
<tr>
<th>Operating System</th>
<th>Ubuntu 16.04 LTS (Xenial Xerus) 64-bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAM</td>
<td>32 GB</td>
</tr>
<tr>
<td>Processor Model</td>
<td>Intel® Xeon® CPU E5-2637 v3 @ 3.50GHz</td>
</tr>
<tr>
<td>Architecture</td>
<td>x86_64</td>
</tr>
<tr>
<td>Socket(s)</td>
<td>2</td>
</tr>
<tr>
<td>Core(s) per socket</td>
<td>4</td>
</tr>
<tr>
<td>Thread(s) per core</td>
<td>2</td>
</tr>
<tr>
<td>NUMA node(s)</td>
<td>2</td>
</tr>
<tr>
<td>L1d cache</td>
<td>32K</td>
</tr>
<tr>
<td>L1i cache</td>
<td>32K</td>
</tr>
<tr>
<td>L2 cache</td>
<td>256K</td>
</tr>
<tr>
<td>L3 cache</td>
<td>15360K</td>
</tr>
<tr>
<td>Parallelization API</td>
<td>OpenMP version 4.0</td>
</tr>
<tr>
<td>Programming Language</td>
<td>C</td>
</tr>
<tr>
<td>Compiler</td>
<td>Intel® C++ version 17.0.2</td>
</tr>
<tr>
<td>Compiler options used</td>
<td>-qopenmp -O3 -xHOST -ipo -std=c11</td>
</tr>
</tbody>
</table>

Table 3.1: Hardware and software specifications.

Since the process of random number generation for simulating Wiener increments is compute-intensive, we first restrict the speedup measurements to the fully deterministic model. We have run the simulation for a $1024 \times 1024$ grid with an initial biomass density (half the maximal value) positioned in the grid cell at the top left corner of the domain.

Figure 3.1 illustrates the amount of gained speedup as more threads are deployed for parallel calculations. One can notice that as the number of threads passes 7, the speedup gradually starts to plateau. Above 10 threads, there is definitely no significant gain in speedup and not only does the deployment of more threads not improve the speedup, but it reduces it ever so slightly. One can readily observe that the parallel configuration can be made approximately 2.3 times as fast as the serial configuration at the highest speedup.

This limited amount of speedup can be attributed to a few factors. The most notable one is the significant overhead due to the creation and subsequent termination of the parallel region inside the while-loop at the beginning of each time step [34, 23, 20, 27]. The parallelization model based on which OpenMP operates is called the *fork/join* model where a master thread branches off into a number of worker threads inside the parallel region. The mentioned overhead originates from this parallelization model since at each time step a number of threads need to be created and scheduled with chunks of loop iterations. Hence, the more the number of time steps, the higher the cost of the parallel region, which is even more
exacerbated in our case because of the small time steps required by the explicit integration methods. The other sources of speedup limitation might originate from the hardware. For instance, as shown in Table 3.1, there are two cores in our hardware setup that are mounted on two separate sockets each. The communications that occur between the cores during the run-time can possibly degrade speedup since the inter-core communication lines are not as fast as the intra-core ones. Furthermore, due to the sheer sizes of the biomass density arrays and carbon concentration arrays, the faster cache lines in the memory hierarchy cannot accommodate them wholly, which means slower cache lines with more memory space have to be used [23]. Another source of inefficiency in parallelization is false sharing that occurs when threads write to their local copies of data in cache lines. Since the local copies must agree with the global copy of the data the writing process has to take place serially. There are techniques such as padding the data arrays [51, 23], which help mitigate the negative effects of false sharing on the speedup.

In our first version of the simulation with the stochastic cell attachment turned on (starting with zero biomass density everywhere except the top left corner, which we kept the same as the deterministic case), we observed a fairly notable deterioration of speedup (see Figure 3.2) that we attribute to the random number generation process. The PRNG, in this case,
has been written in the serial part of the while-loop where at the beginning of each time step, a for-loop would generate a Gaussian random number (Wiener increments in the Euler-Maruyama scheme) for each grid cell using pseudo random number generating algorithms. As it turns out this for-loop imposes a great computational cost to each time iteration. One could ascribe the actual reason for this parallelization inefficiency to implications of Amdahl’s Law [1]. The law states that when a fraction $f$ of a computation’s time is subjected to parallelization with $n$ processors and the rest, $(1 - f)$, remains serial, the speedup is given by

$$speedup(f, n) = \frac{1}{(1 - f) + \frac{f}{n}}. \quad (3.4)$$

Obviously, the speedup is impacted negatively by a huge computational cost when the PRNG is implemented in serial. In view of (3.4), this means that $f \ll 1$ or equivalently $(1 - f) \approx 1$. Logically, we need to parallelize the PRNG algorithm in order to recover, at the least, the speedup trend of the deterministic case. Figure 3.3 shows the speedup for the while-loop after the PRNG is parallelized as well. We describe the details of the algorithm used for this purpose later. As can be seen evidently, not only is the parallel speedup recovered but also a higher speedup factor is accomplished (as big as 5 times at the maximum). Although a large number of computations are added inside each time step
in comparison with the deterministic no-attachment case, the inclusion of the PRNG in
the parallel region of the stochastic cell attachment simulation has turned out to be more
effective. From (3.4) one can verify that using a parallel PRNG reduces the time fraction,
\((1 - f)\), spent in the serial portion and adds to the time fraction, \(f\), spent in the parallel
portion, which results in a larger speedup. This observation is corroborated by the literature
on the speedup in a multicore system [24].

We address why better speedup results are accomplished for the parallel-PRNG simula-
tions by comparing inherent algorithmic differences. The typical average time in seconds for
the while-loop to complete in the deterministic no-attachment case is between 2 and 5 where
the former is achieved by 10 threads and the latter by 1 thread. For the parallel-PRNG
stochastic attachment case the numbers are 11 and 55 on average where 11 corresponds to
16 threads and 55 to 1 thread. Evidently, the process of random number generation for
implementing the Euler-Maruyama scheme adds more overhead to the simulation. However,
as one can verify by checking Figure 3.1 and Figure 3.3, the accomplished speedup for the
stochastic case is approximately 2.2 times greater than that of the deterministic case at the
maximum values. We attribute this behavior to the differences that exist between the for-
loops that handle random number generation for each grid cell and the for-loops that handle

Figure 3.3: The speedup as a function of number of threads for parallel PRNG. Calculated
for 2 units of simulation time.
the numerical integrations. In the latter case we have to impose boundary conditions when boundary cells are being processed. Dealing with these special cases would definitely diminish from the efficiency of the code and is possibly the bottleneck for the parallel speedup. Such source of inefficiency does not arise in the case of the random number generating for-loops. There are no special cases, which means we can easily collapse the for-loops over \( i \) and \( j \) into a single one. Also the PRNG algorithm generates two random numbers simultaneously, which means we only need to iterate half the total number of grid cells. These characteristics, in our opinion, make these for-loops more responsive to parallelization, and hence a higher speed up factor is obtained as we add more threads to the parallel region.

### 3.3 Random Number Generation

As has been pointed out a few times so far, the Euler-Maruyama scheme (3.2) requires the Wiener increments, \( \delta W_n \), for the numerical construction of the attachment factor \( \phi \). In order to generate the Wiener increments for time integration inside each time step, we use a combination of PRNGs.

We recall that the Wiener increments are independent Gaussian random variables with \( E(\delta W_n) = 0 \) and \( Var(\delta W_n) = h \). We can sample random numbers with these properties from the standard normal distribution, \( \mathcal{N}(0,1) \), scaled by the square root of the time step size \( \sqrt{h} \mathcal{N}(0,1) \). We have

\[
E \left( \sqrt{h} \mathcal{N}(0,1) \right) = \sqrt{h} E \left( \mathcal{N}(0,1) \right) = 0,
\]

and

\[
Var \left( \sqrt{h} \mathcal{N}(0,1) \right) = h Var \left( \mathcal{N}(0,1) \right) = h.
\]

Thus the properties of the sampled numbers are those required for numerically simulating Wiener increments. There are a few algorithms for generating \( \mathcal{N}(0,1) \) random numbers. One of the well-known algorithms is called the Box-Muller method, where using two independent random variables with uniform distribution on the unit interval \((0,1)\), or in short notation \( U(0,1) \), two independent random variables with standard normal distribution are constructed \([3]\). The formulae for this method are based on trigonometric functions and the natural logarithm. As it turns out, trigonometric functions are computationally expensive. The Marsaglia polar method eliminates the trigonometric functions by using polar transformations and is deemed to be more efficient \([40]\). Using this method, the pseudo random
numbers are given by
\[ Z_1 = V_1 \sqrt{-2 \ln W/W}, \]
\[ Z_2 = V_2 \sqrt{-2 \ln W/W}, \] (3.5)

where \( Z_1 \) and \( Z_2 \) are two independent normally distributed random numbers, \( V_1 = 2U_1 - 1 \) and \( V_2 = 2U_2 - 1 \) with \( U_1 \) and \( U_2 \) being two \( U(0, 1) \) random numbers and \( W = V_1^2 + V_2^2 \leq 1 \).

The computational implementation of the above method requires generating \( U_1 \) and \( U_2 \) continuously until \( W \) falls within \((0, 1)\). Although some of the generated \( U(0, 1) \) numbers are discarded, this method is oftentimes more efficient than the Box-Muller method especially when a large batch of random numbers are to be generated for an application [30].

The dependence of the Marsaglia polar method on the \( U(0, 1) \) random numbers requires another algorithm for generating pseudo random numbers in the unit interval \((0, 1)\). We have implemented the corresponding PRNG by utilizing a combination of the linear congruential pseudo random number generators (LCGs) and Lagged Fibonacci generators (LFGs) to generate random integers that are subsequently mapped to \( U(0, 1) \) random numbers.

An LCG generates integer pseudo random numbers recursively, namely
\[ R_{n+1} = (a \cdot R_n + c) \mod m, \] (3.6)

where \( R \) stands for the integer pseudo random numbers, \( a \) is called the multiplier, \( c \) the increment and \( m \) the modulus. For initializing, an integer initial value or seed, \( R_0 \), is replaced in (3.6) and subsequent random numbers are generated recursively. As a consequence of the existing modulus operation the generated integers assume values between 0 and \( m - 1 \) [30].

Considering the sheer number of random numbers needed for a typical simulation of (2.38), the deployed LCGs should have long periods; that is, these should produce sufficiently long sequences of random numbers before the same sequence is repeated. As a result, we have chosen LCGs with \( m := 2^p = 2^{63} \) and other characterizing parameters \((a, c)\) provided in [38] (see Table 4 therein) so that longer periods are achieved.

In order to attain even longer periods, one can use LFGs in conjunction with LCGs [30]. These generators have the general form
\[ R_n = R_{n-r} \star R_{n-s}, \] (3.7)

where \( r \) and \( s \) are the lags with \( 0 < s < r \), \( n \geq r \) and \( \star \) is a binary operator, e.g. addition \((\mod m)\), subtraction \((\mod m)\), etc. We have used this method with an addition \((\mod m)\)
binary operator for generating random integers. The computational implementation of the LFG requires an initialization sequence of random integers with length $r$. This sequence is constructed by an LCG [30]. Using this method, a maximal period of $2^{r-1}(2^r - 1)$ can be achieved if at least one member of the initialization sequence is odd and suitable values are chosen for the lags [41, 39, 43]. We have acquired our lag values, $r = 127$ and $s = 97$, from [4] (see Table 1 therein).

The procedure for generating the required Wiener increments for implementing the calculations consists of the following steps inside each time iteration until the time loop terminates:

1. Every time iteration is started by generating two random integer sequences whose lengths are half the number of grid cells (we work with even numbers here). We use two distinct LCGs\(^2\) – designated as LCG2 and LCG3 – to generate the sequences i.e. each LCG is responsible for one of them. These sequences are stored in two separate arrays\(^3\).

2. Using another LCG – labeled LCG1 – two initialization sequences of random integers (needed for running the LFG) are constructed for every grid cell and stored in arrays of size $r$: one array construction is seeded by each element of the sequence generated by LCG2 and the other one by each element of the sequence generated by LCG3. It should be noted that each element of the seeding sequences is in fact associated with a grid cell.

3. The LFG algorithm uses the generated initialization sequences and produces two sequences of integer random numbers, which are stored in arrays of size $r - s$. Dividing the random integers in each sequence by the modulus results in random numbers that are between 0 and 1 i.e. numbers calculated as

   \[ U_n = \frac{R_n}{m}, \]  

   \[ (3.8) \]

   would seem to be $U(0, 1)$ [30]. The final product of this step is two arrays of $U(0, 1)$ random numbers.

\(^2\)Using two LCGs (LCG2 and LCG3) instead of one is not necessary. Our main intention for implementing two LCGs is to presumably minimize the impact of any possible statistical correlations that might exist in a sequence generated by just one LCG.

\(^3\)The main reason for storage in arrays is to simplify the incorporation of this portion of the code in the parallel region of the multithreaded version. The serial version of the same portion of the code can be written by non-array variables as well.
4. The Marsaglia polar method is executed within a for-loop that is iterated half the number of grid cells where within each iteration the two arrays of \( U(0, 1) \) random numbers are checked for elements (via a while-loop) that satisfy the required conditions\(^4\). The picked element from the first array is used to calculate \( V_1 \) and the picked element from the second one is used for computing \( V_2 \) (see Eq. (3.5)). The resulting \( \mathcal{N}(0, 1) \) pair of random numbers are used to fill up an array that has the same size as the total number of grid cells. The components of this array are used later for calculating the Wiener increments associated with each grid cell.

This particular approach for generating the Wiener increments using the described combination of PRNGs is extremely empirical. It is primarily based on our observations of the speed, efficiency and practicality of the whole random number generating process in the context of the numerical treatment of our proposed model. We have taken the precaution to examine the simulation results for any signs of either a specific or recurring pattern of random cell attachments. Based on our observations we conclude that the above procedure for generating \( \mathcal{N}(0, 1) \) numbers is adequate for our simulation purposes.

### 3.4 Random Number Generation in Parallel

We have presented the speedup results for the serial and parallel PRNGs in Figure 3.2 and Figure 3.3. We notice that the parallelization of the PRNG saves a sizable amount of computation time whereas generating random numbers in serial reduces the speedup in a way that renders the whole parallel computing somewhat pointless. In our attempt to parallelize the random number generation process we have utilized different approaches for different component parts of the PRNG.

In order to make the two main seeding sequences (generated by LCG2 and LCG3) be amenable to implementation within the parallel region, we have applied the fundamentals of a technique called the leapfrog method to the elements of these sequences in each time step. We should clarify that this is accomplished by adapting this method to the problem at hand and the following explanation, which uses the idea of threads in a multiprocessor system, is only given here to familiarize the reader with the basics. The exact details of how we have used it are laid out after the following short introduction.

\(^4\)Since the while-loop that checks for the required conditions has never become infinite in our simulations, we assume that each time the two arrays of \( U(0, 1) \) random numbers are examined, a suitable pair is found.
The basic idea behind the leapfrog method is to produce the same sequence of random numbers by using a multiprocessor system where each processor generates a non-overlapping subsequence, which is a subset of the main sequence. One must be able to reconstruct the main sequence by stringing together the subsequences independent of the number of processors (threads) incorporated inside the parallel region of the code [6]. The leapfrog method accomplishes this task by assigning each processor a subsequence of the original sequence whose members are separated by the number of threads in the parallel region. That is the main reason this technique is called the leapfrog method since the threads leapfrog by the number of threads over the original sequence: assuming there are \( P \) threads in total and that thread number \( T \) is one of them, then according to the leapfrog method the subsequence \( R_T, R_{T+P}, R_{T+2P}, R_{T+3P}, \ldots \) is exclusively produced by thread number \( T \) [6].

We have deployed the leapfrog method in a slightly different manner, which is more akin to what is described in [15]: assuming there are \( N \) branches in total and that branch number \( B \) is one of them, then the subsequence \( R_B, R_{B+N}, R_{B+2N}, R_{B+3N}, \ldots \) is exclusively produced by branch number \( B \). In the context of our model the branches are replaced by grid cells. Instead of leapfrogging over the main sequences of random integers by the number of threads, the jump is made by half the number of total grid cells (the number of random integers in each sequence generated by LCG2 and LCG3). Here it does not matter which thread is using a specific element of the sequences generated by LCG2 and LCG3 to produce random integers for the next time step. The main goal is to assign a specific grid cell with random integers that are half the number of grid cells away from one another in the main sequence of possible random numbers. Without implementing the leapfrog method, the grid cells would only experience a single shift in the assigned values of random integers \( i.e. \) the assigned random integer to the \( (n) \)th cell for the next time step would be the value assigned to the \( (n + 1) \)th cell for the current time step.

No matter what number of grid cells is assumed in the beginning of the simulation, the same sequence of random numbers are generated for the same starting seed. Obviously for the same amount of simulation time the random numbers generated for the coarsest grid are a subset of the random numbers generated for the finest grid.

The multiplier \( a \) and the increment \( c \) in (3.6) must be modified for the implementation of leapfrogging in the described manner. The modified parameters are given by [15]

\[
A_N = a^N \mod m,
\] (3.9)
and

\[ C_N = c \frac{A_N - 1}{a - 1} \mod m, \quad (3.10) \]

where \( N \) is half the number of grid cells. In the actual simulation we first run LCG2 and LCG3 recursively to fill up the arrays that store the random integers and then using the modified version of (3.6), namely

\[ R_{n+N} = (A_N \cdot R_n + C_N) \mod m, \quad (3.11) \]

inside a parallel for-loop, the spawned threads generate the next random integers in the sequence during each time step. The most prominent feature of leapfrogging in the parallel version is that the overall sequence of random integers is exactly the same as the one generated by (3.6) in the serial version of the code.

The remaining step in numerical simulation of the Wiener increments is parallelizing the Marsaglia polar method. This turns out to be trivial as we only need to execute a parallel for-loop that uses the random integers generated by leapfrogging and parallel computing and feeds them to the Marsaglia polar method to produce a pair of \( \mathcal{N}(0,1) \) random numbers as was expounded in the previous section. An array with a size equal to the total number of grid cells is used to store the \( \mathcal{N}(0,1) \) random numbers as before. Subsequently during the numerical evaluation of (3.2) these numbers are multiplied by the square root of the time step size to produce the Weiner increments.

### 3.5 Numerical Stability and Time Step Size

The numerical simulation of the system of SDEs (2.38) is very sensitive to the time step size since the explicit method of integration, namely the Euler method, stays stable within a specific stability region [18]. Consequently, the time step sizes have to be chosen very small for the numerical treatment of nonlinear diffusion PDEs with the diffusion coefficient (2.3) especially when the biomass density approaches its maximal value [21].

Our choice of Euler method, which is the lowest order time integration scheme, is for the most part because of our assumption that the biomass solutions of (2.38) are not smooth enough (higher order derivatives do not exist) for implementation of higher order ODE solvers. We know that the stochastic attachment of cells to random sites on the substratum introduces stochastic fluctuations in the biomass density solutions. As a result of these fluctuations the smoothness of the solutions cannot be guaranteed. We have also not been
able to gain much from applying higher order SDE solvers, such as Milstein method [30], since the higher order terms in the numerical scheme for these solvers depend on explicit derivatives with respect to the dependent variable. For the current form of the AF SDE, (2.35), all these higher order terms vanish and the numerical scheme becomes equivalent to (3.2).

To avoid the numerical instability issues, we have implemented the time integration of (2.38) with a time step size calculated by a criterion akin to the well-known Courant-Friedrichs-Lewy (CFL) condition [7]. This condition is necessary for the convergence of a numerical approximation of a PDE (see Theorem 4.5 in [52]). In the particular case of parabolic PDEs, if the explicit numerical approximation for the grid cell length $\Delta x$ is convergent, then the time step, $h$, has to be $h = o(\Delta x)$ as $h \to 0$ (see Theorem 4.7 in [52]). In a 2-dimensional setting, the CFL condition reads

$$C \equiv \frac{u_x h}{\Delta x} + \frac{u_y h}{\Delta y} \leq C_{max},$$  \hspace{1cm} (3.12)

where $C$ is called the Courant number, $u_x$ and $u_y$ are the speeds by which the exact solution (in our case the biomass density component of the solution) traverses the domain, $\Delta x$ and $\Delta y$ are the length of the grid cells in $x$ and $y$ directions and $h$ is the time step size.

We can derive an upper bound for the time step size by applying (3.12) to our diffusive nonlinear problem. For this, we assume the spatial discretization is the same in both directions and equal to $N$ and that the domain sides are equal to $L$; thus, the grid cell lengths are $\Delta x = \Delta y = \frac{L}{N}$. Furthermore, by symmetry, the speeds $u_x$ and $u_y$ are equal and given by $D(M)/N$ where $D(M)$ is the density dependent diffusion coefficient (2.3). Therefore, the upper bound is given by

$$h \leq \frac{1}{2D(M)} \left( \frac{L}{N} \right)^2.$$ \hspace{1cm} (3.13)

To derive the above relation, we have assumed $C_{max} = 1$, which is a valid assumption for the explicit method used in our simulations [32, 52].

The upper bound for the time step size is inversely proportional to the diffusion coefficient. Since the diffusion coefficient is a function of the biomass density, the upper bound cannot be determined by the model parameters solely. In our computer simulations, we typically use biomass density values 2-4 percent below the maximal attainable value and compute the upper bound for the time step size (3.13) by replacing our estimate in the diffusion coefficient. This tactic has often preserved the stability of the numerical method where
the final goal is running a single simulation with particular parameters. However, this is not guaranteed and the numerical instability is manifested as soon as the initial data are changed, the grid is refined or other model parameters are adjusted. This issue becomes especially cumbersome when computer experiments involve studying the stochastic cell attachments that are simulated by several different starting seeds and/or parameter sensitivity analyses are performed. To address the shortcomings of our tactic, we have incorporated an adaptive time step size calculation in the sense that for each future time step a new step size is computed by the current highest attained value of the biomass density. In order to implement this strategy in our computer code, we first initialize the time step size to a sufficiently small value: arbitrarily pick a biomass density value \( e.g. \) 6 percent below the maximal value. This step size value is stored as the upper bound in a variable. After using the initial step size to approximate the current biomass density, carbon concentration and attachment factor (see Algorithm 1), a for-loop is executed that loops through the biomass density array and calculates the time step size associated with each grid cell. Subsequently, using a simple minimum-returning function, a variable that stores the smallest step size is updated in each iteration of the for-loop, and the final value of this variable, if less than the upper bound, is used for the next time step. This procedure is iterated until the simulation is terminated.

It should be noted that the aforementioned for-loop can be executed in parallel by using OpenMP work-sharing construct \#pragma omp for reduction(min:) \[48\]. Hence, one can reduce the serial calculations that are introduced due to the added process and perform them in the parallel region of the code.
Chapter 4

Simulation Results and Computational Analysis

Having a complete mathematical model for the stochastic attachment of bacterial cells in the context of cellulolytic biofilms and a computational framework that enables us to explore the model by computer simulations, we proceed to check whether two-dimensional simulations of the model show the expected features of cellulolytic biofilm formations. Later, we investigate the model in a one-dimensional setting (for computational cost considerations) in order to draw some important conclusions on the numerical convergence of state variables, $M$ and $C$, due to grid refinement and the relation between the grid refinement and stochastic fluctuations. Sensitivity analysis results are obtained for the two-dimensional version of the model and are discussed thoroughly in the last section.

4.1 2D Simulations

We begin by showcasing a few simulation examples in the 2D setting. These examples consist of two distinct cases with two different rates of stochastic cell attachment. For the first case, the simulation is initiated with a clean domain, i.e. the initial data for biomass density equals zero everywhere in the domain. We change this initial state in the second case by initiating the simulation with a small established colony at the center of the domain. Our goal is to investigate qualitatively the effect of an established colony on the random attachment of free cells and the ensuing colony formations and developments. In both of these scenarios, we look into variations in the biofilm development that might arise from changing the rates of bacterial cell attachments.
We have also examined a third case where the stochastic attachments exhibit a particular pattern along the spatial dimensions. The idea behind this case is to simulate an environment where the aqueous phase flows over the substratum and transports the free cells or in other words the free cells show a directional motion inside the domain. If the floating free cells enter the domain with an approximately constant flux and we assume that the free cells are capable of attachment to the substratum from the very point of entrance\(^1\), then expecting more instances of random attachments upstream and less so downstream would be justified within the current framework. We have accomplished simulating this scenario by manipulating the initial data for the AF in a special manner.

The simulation parameters for all these cases (note that the initial data for the AF used in the third case will be discussed in the corresponding section) are given in Table 4.1.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motility Coefficient</td>
<td>(\delta)</td>
<td>(1.0 \times 10^{-6})</td>
<td>[12]</td>
</tr>
<tr>
<td>Diffusion Coefficient Exponent 1</td>
<td>(\alpha)</td>
<td>4.0</td>
<td>[12]</td>
</tr>
<tr>
<td>Diffusion Coefficient Exponent 2</td>
<td>(\beta)</td>
<td>4.0</td>
<td>[12]</td>
</tr>
<tr>
<td>Maximum Consumption Rate</td>
<td>(\tilde{\Upsilon})</td>
<td>0.4</td>
<td>[12]</td>
</tr>
<tr>
<td>Half Saturation Concentration</td>
<td>(\tilde{\kappa})</td>
<td>0.01</td>
<td>[12]</td>
</tr>
<tr>
<td>Cell Loss Rate</td>
<td>(\tilde{\lambda})</td>
<td>0.42</td>
<td>[12]</td>
</tr>
<tr>
<td>Control Parameter of (f(\phi))</td>
<td>(\sigma)</td>
<td>0.00001</td>
<td>assumed</td>
</tr>
<tr>
<td>Shift Parameters of (f(\phi))</td>
<td>(a &amp; b)</td>
<td>0.99 &amp; 1.00</td>
<td>assumed</td>
</tr>
<tr>
<td>Density of an Attached Cell per Time Step</td>
<td>(\gamma_1)</td>
<td>0.01</td>
<td>assumed</td>
</tr>
<tr>
<td>Biomass Attachment Exponent 1</td>
<td>(\gamma_2)</td>
<td>1.0</td>
<td>assumed</td>
</tr>
<tr>
<td>Biomass Attachment Exponent 2</td>
<td>(\gamma_3)</td>
<td>1.0</td>
<td>assumed</td>
</tr>
<tr>
<td>Constant Coefficient in the AF SDE</td>
<td>(\tilde{\psi})</td>
<td>0.04 &amp; 0.08</td>
<td>assumed</td>
</tr>
<tr>
<td>Exponent 1 in AF SDE</td>
<td>(\nu_1)</td>
<td>1.0</td>
<td>assumed</td>
</tr>
<tr>
<td>Exponent 2 in AF SDE</td>
<td>(\nu_2)</td>
<td>1.0</td>
<td>assumed</td>
</tr>
<tr>
<td>Initial Data for (\phi)</td>
<td>(\phi_0)</td>
<td>0.0</td>
<td>assumed</td>
</tr>
<tr>
<td>Starting Integer Seed for the PRNG</td>
<td>(sd)</td>
<td>10</td>
<td>assumed</td>
</tr>
<tr>
<td>Domain Length</td>
<td>(L)</td>
<td>1.0</td>
<td>assumed</td>
</tr>
<tr>
<td>Domain Width</td>
<td>(W)</td>
<td>1.0</td>
<td>assumed</td>
</tr>
<tr>
<td>Grid Size</td>
<td>(N \times M)</td>
<td>(2^9 \times 2^9)</td>
<td>assumed</td>
</tr>
</tbody>
</table>

Table 4.1: Model parameters used for 2D simulations.

\(^1\)We are also implicitly assuming that the time scale of the transport speed is relatively slower than that of the biofilm formation, which would be commensurate with the simulation results shown later.
4.1.1 Case 1: Initially Clean Domain

This case has been simulated with zero initial biomass inside the domain. The carbon concentration is set to its maximal value over the entire domain\(^2\). We have controlled the emerging characteristics of these simulations, namely the rate of stochastic cell attachment and the overall biofilm formation features, by adjusting the constant coefficient in the AF SDE (2.35), \(\psi\). Simulation 2 has been run with a \(\psi\) value that is half of the one used for simulation 1 (see Table 4.1). By checking Figure 4.1 (simulation 1) and Figure 4.2 (simulation 2), we can easily verify that the biomass density equals zero and the carbon concentration equals 1 everywhere in the domain at \(t = 0\) (the plotted time points for these figures are not the same after \(t = 0\)). As the simulations progress temporally, small colonies start to emerge gradually. Figure 4.1 demonstrates that the seeds of the initial colonies begin to become visible around \(t = 25\). These colonies grow radially and eventually manifest crater-like features (see Figure 1.2) resembling those of the cellulolytic biofilms (see Figure 1.3). As the radial growth continues, some colonies coalesce and form communities. In turn, these communities grow and join other communities, which produce even larger communities.

The morphological features of biomass development are reflected in the carbon concentration figures and one can easily spot the outline of either individual colonies or large communities carved out in the substratum, which is a consequence of carbon depletion as bacteria consume it for growth and proliferation.

By comparing Figure 4.1 and Figure 4.2, one can observe some variations in colony and community development. The stochastic cell attachments occur rather later (around \(t = 45\)) and more scarcely for simulation 2. As a consequence, the initial colony formations are patchier and it takes a bit longer for these colonies to merge and form communities. Unlike simulation 1, the substrate supply lasts longer to around \(t = 80\) in simulation 2 to the point that some remaining spots of the substratum that still contain maximal carbon concentrations have become grounds for small starting colonies.

The smaller \(\psi\) values cause the amplitude of the variations in the AF to be less pronounced (see Figure 4.3); therefore, it becomes less likely for the AF values, if any, to fall within the prescribed effective domain of the impulse function. In other words, we decrease the rate of stochastic cell attachment events by reducing the magnitude of the stochastic change in the AF. Thus during simulation 2, only a few scattered colonies form, and in between the attachment events the domain becomes less available to new cell attachments since the

\(^2\)This is the case for all simulations in this chapter.
growth and consumption rates for biomass and substrate are effectively faster than that of the attachment events.

We have plotted the fraction of cells whose $\phi$ value is within the support of the impulse function, $f(\phi)$. As we can see in Figure 4.4 this fraction, denoted by $F$, grows and fluctuates as a function of time for both $\psi$ values. The rate of growth of $F$ is higher in the case of greater $\psi$. Again, this can be attributed to the more pronounced amplitude of stochastic fluctuations in this case, which causes a greater number of cells to become viable for random attachments earlier in time and before the biomass growth and carbon consumption renders most of the domain inhospitable to additional random attachments. One can also observe that in the case of greater $\psi$ the fraction becomes constant towards the end of simulation. These are the cells whose $\phi$ value remains within the support of $f(\phi)$ after a steady state is attained. As we explained in Chapter 2 the effect of the constant impulse from these cells is nullified by the coefficient $\gamma_1(1 - M_{i,j})^{\gamma_2}C_{i,j}^{\gamma_3}$ in (2.34) where for these cells $C_{i,j}$ is zero and thus, their unwanted effect gets eliminated.
Figure 4.1: Simulation of (2.38) with initially $M = 0$ everywhere and $\psi = 0.08$. The biomass densities (left) with their corresponding carbon concentrations (right) are provided for different time points.
Figure 4.1: Simulation of (2.38) with initially $M = 0$ everywhere and $\psi = 0.08$. The biomass densities (left) with their corresponding carbon concentrations (right) are provided for different time points.
Figure 4.2: Simulation of (2.38) with initially $M = 0$ everywhere and $\psi = 0.04$. The biomass densities (left) with their corresponding carbon concentrations (right) are provided for different time points.
$t = 70$

$t = 80$

Figure 4.2: Simulation of (2.38) with initially $M = 0$ everywhere and $\psi = 0.04$. The biomass densities (left) with their corresponding carbon concentrations (right) are provided for different time points.
Figure 4.3: Here we have plotted two sample paths with different $\psi$ values for one component of the AF, $\phi$, vs. time. This component is located close to the right boundary and middle of the domain, which in this case is the 230729th entry. One can easily verify that the overall amplitude of stochastic changes in the line-circle path does not exceed that of the line-asterisk path during the whole time. Also, it takes longer for the line-circle path to reach steady state since the attachment event in this case takes place later in time.

Figure 4.4: The change in the fraction of cells whose $\phi$ value is within the support of $f(\phi)$ vs. time. The fractions have been recorded at the end of 0.1 intervals in time.
4.1.2 Case 2: Initially Colonized Domain

Here we have simulated a scenario where at $t = 0$ a small bacterial colony is centered in the middle of the domain. The main goal for this setup is to examine the effect of stochastic cell attachments on the growth of the central colony and vice versa.

As demonstrated by Figure 4.5 and Figure 4.6, the central colony grows radially with a distinct circular front that sweeps through the domain. In the wake of this front, the substrate is consumed, which eventually renders the area within the central crater-like structure unavailable to random cell attachments.

As before, by changing the value of $\psi$ the frequency of attachment events can be controlled. For higher values of this parameter, more random attachments occur over the areas that are still replete with substrate and are not occupied by biomass. This is evidently the case in Figure 4.5 where by $t = 35$ numerous small colonies have formed and many new cell attachments are visible. The growth rate for these small peripheral colonies is such that by $t = 40$ many of them have merged and developed into communities. Subsequently the biofilm communities surround the central colony from all directions and merge into it. This interaction distorts the sharp front of the central colony and suppresses its further growth toward the boundaries of the domain. The situation is very different when $\psi$ is set to half the previous value. Figure 4.6 displays how the lower rate of random attachments can emphasize the usual characteristics that we expect from a deterministic model of a single-colony cellulolytic biofilm.

In the previous section, we discussed how a lower $\psi$ value practically decreases the overall number of attachment events both spatially and temporally in a single simulation. Obviously when there are less random cell attachments, the central colony has more substrate available to it for growth. As we can see, the first telltale signs of attached cells become visible around $t = 45$. However, these scattered colony seeds are too few to distort the growth of the central colony; thus, the circular front reaches the boundaries unimpeded by any major peripheral colonies that often form with higher $\psi$ values. The very few peripheral colonies that find the chance to form are able to slightly contort the sharp front but are merged in and swept over without leaving a significant trace or distortion behind.
Figure 4.5: Simulation of (2.38) with initially $M \neq 0$ only in the center and $\psi = 0.08$. The biomass densities (left) with their corresponding carbon concentrations (right) are provided for different time points.
$t = 40$

$M$

$t = 45$

Figure 4.5: Simulation of (2.38) with initially $M \neq 0$ only in the center and $\psi = 0.08$. The biomass densities (left) with their corresponding carbon concentrations (right) are provided for different time points.
Figure 4.6: Simulation of (2.38) with initially $M \neq 0$ only in the center and $\psi = 0.04$. The biomass densities (left) with their corresponding carbon concentrations (right) are provided for different time points.
Figure 4.6: Simulation of (2.38) with initially $M \neq 0$ only in the center and $\psi = 0.04$. The biomass densities (left) with their corresponding carbon concentrations (right) are provided for different time points.

4.1.3 Case 3: Spatial Variation in Random Attachments

This case is an instance of possible scenarios that the current modeling approach allows us to explore. The main purpose here is to simulate a situation where the random attachments occur with varied spatial rates over different parts of the domain as if the aqueous phase were flowing from one domain boundary to the other bringing in free cells as it entered the domain. To simulate the described scenario, we have assumed that the free cells enter the domain from the right side. We make no assumptions on the actual numerical values of number of cells or the transport speed since these do not appear in the current formulation of the model. The sole purpose here is to see whether the mentioned scenario could be
produced by manipulating the model parameters or the initial data.

This case is somewhat similar to case 1 where we have a clean domain. The main difference is the initial data used for the AF denoted by $\phi_0$. We have initialized the AF to zero everywhere for simulating case 1. As a result the underlying mechanism for random attachments have been initially started off on the same footing over the whole domain. As it turns out, no particular spatial variation has been displayed for the random attachments (see Figure 4.2 and Figure 4.1) and the whole domain has experienced equal chance for accommodating these events especially in the very beginning of the simulations. We have realized that the spatial variation can be introduced by using a special treatment of $\phi_0$ where instead of initializing it equally, it can be set up with a monotonically increasing initial data that grows in value from left to right of the domain. Consequently, the grid cells closer to the right boundary start from a more amenable situation for attachments since their initial AF value is closer to the prescribed effective domain of the impulse function, which gives them a higher chance of falling within that domain.

Following the above arguments, we have initialized $\phi$ as such: the leftmost column of grid cells is initialized to zero and as the initialization process moves toward the right end columns the value of the initial data used for the next column is incremented by 0.0005 (in the case of a 512 × 512 grid, the rightmost column is initialized to 0.2555). The simulation results (see Figure 4.7) show a definite spatial bias for more attachments near the right boundary. The exhibited features can also be thought of as a situation where the free cells show more inclination to attachment over some parts of the substratum that are more hospitable to colonization i.e. this behavior does not necessarily and solely arise from the scenario we laid out in the opening of this section.

In terms of colony formation and substrate consumption, evidently the first major colonies form over the right side of the domain; therefore, the substrate is consumed according to the same pattern of biofilm growth. These initial colonies merge into one another and form a larger community that sweeps from right to left. As the substratum degrades due to biofilm growth, the amount of space for new random attachments dwindles and hence the main community takes over most of the domain and any new small colonies are absorbed by it. The observed mode of biofilm formation produces a directional attitude in biofilm growth, substrate consumption and substratum degradation even if its driving mechanism is fueled by a stochastic process. This is in contrast to the disorderly random behavior one expects to observe as before, e.g. case 1.
Figure 4.7: Simulation of (2.38) with initially $M = 0$ everywhere and $\psi = 0.04$. We assume that there is an inflow of free cells from right to left. The biomass densities (left) with their corresponding carbon concentrations (right) are provided for a few time points.
4.2 Grid Refinement Studies

We have performed grid refinement studies on the SDE system (2.38) to investigate the effect of several different grid sizes on the temporal behavior of the biomass density and the carbon concentration. Because of high computational costs incurred by 2D simulations, we have opted for a 1D setting in these studies.

Due to the stochastic nature of the problem, the simulations have been executed (sampled from possible outcomes) for a number of times by initializing the PRNG seeds differently each time. We have implemented a statistical treatment of the obtained results for the biomass density and the carbon concentration. The statistical approach involves calculating a few relevant statistics, including the sample mean, the sample variance, the sample standard deviation and the standard error of the sample mean.

The effects of grid refinement are evaluated in a couple of different manners:

1. Studying the convergence of the sample means and the standard error of the sample means for spatially integrated biomass density and carbon concentration.

2. Investigating the degree of stochastic fluctuations as the attachment rate is increased by refining the grid.

In order to familiarize the reader with the biofilm growth and the substrate consumption characteristics in a 1D setting, some examples of the simulations that are the target of grid refinement studies are shown below. It should be noted that the temporal dimension is introduced along the vertical axis so that a 2D color map can be produced.

The simulation parameters used for all 1D simulations are mostly the same as the ones provided in Table 4.1. The few different ones are given in Table 4.2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant Coefficient in the AF SDE</td>
<td>$\psi$</td>
<td>0.1 &amp; 0.2</td>
<td>assumed</td>
</tr>
<tr>
<td>Starting integer seed for the PRNG</td>
<td>$sd$</td>
<td>1-100</td>
<td>assumed</td>
</tr>
<tr>
<td>Grid size</td>
<td>$N$</td>
<td>$2^4$-$2^{13}$</td>
<td>assumed</td>
</tr>
</tbody>
</table>

Table 4.2: Model parameters that are different in 1D simulations.

We start by showing some simulation results for a case where the spatial length of random attachments is $\frac{1}{16}$ of the domain length (the manipulation of the attachment size is elaborated in Section 4.2.1; the idea is to have the same spatial sizes for attachments independent of
Figure 4.8: 1D Simulation of (2.38). Simulation time is shown along the vertical axis. The color map on the left represents the temporal behavior of biomass in a 1D domain and the one on the right indicates the corresponding substrate consumption. The spatial size of each random attachment is $\frac{1}{16}$ of the domain length.

We have chosen the value of $\psi$ such that only a few stochastic cell attachments occur during the simulation. For instance, in simulations represented by Figure 4.8, the number of random attachments that are sufficiently visible is four, but upon further examination, one can spot two more attachment instances that are too close to the existing colonies to be distinguishable. Thus, the total number of random cell attachments is six.

The next case exhibits a moderately different time evolution of biomass and substrate. Here, the spatial length of random attachments is $\frac{1}{128}$ of the domain length. In order to maintain a relatively low rate for the stochastic attachments so that the entire domain is not overpopulated by the sheer number of associated small colonies, we have lowered the value of $\psi$ in this case. As a result, the total number of random attachments during the simulation time (see Figure 4.9) is around 17. Although there is an obvious increase in the total number of random attachments relative to the previous case, the dissimilarities appearing in the pattern of biomass growth and substrate consumption are not just a consequence of this change. It is presumably due to the eight-fold reduction in the spatial size of random attachments as well.

The final case for which we have provided some visualizations of simulation results, is based on a very different scenario where the number of stochastic cell attachments is proportional to the number of grid cells. It should be noted that $\psi$ is kept the same for all simulations in this case.

$$\psi = 0.2, \quad sd = 15, \quad N = 4096$$
Figure 4.9: 1D Simulation of (2.38). Simulation time is shown along the vertical axis. The color map on the left represents the temporal behavior of biomass in a 1D domain and the one on the right indicates the corresponding substrate consumption. The spatial size of each random attachment is $\frac{1}{128}$ of the domain length.

The most prominent observation one can make examining Figure 4.10 is the formation of a fairly narrow short-lived biofilm community along the time axis for the most refined grid. This feature can be attributed to a couple of factors, including the overall higher number of random attachments and the close proximity of these attachments, which result in the formation of numerous small colonies very close to one another that in a short amount of time merge and form a relatively homogeneous community. This mode of biofilm growth is also reflected in the substrate plots where we observe a rise in the number of sharp features that dominate the consumption pattern and subsequently a more evenly-shaped disappearance of the substrate in time.
ψ = 0.2, $sd = 15$, $N = 256$

ψ = 0.2, $sd = 15$, $N = 1024$

ψ = 0.2, $sd = 15$, $N = 4096$

Figure 4.10: 1D Simulation of (2.38). Simulation time is shown along the vertical axis. The color map on the left represents the temporal behavior of biomass in a 1D domain and the one on the right indicates the corresponding substrate consumption. Utilization of finer grids for computer simulations entails more random attachments in this case.
4.2.1 Refinement and Convergence of Solution Components

One of the basic tests in the context of numerical analysis is to check whether the solution of a problem (assuming its existence and uniqueness) converges to a certain value or set of values as the computational precision is enhanced. In the present case, where we utilize a grid to divide the spatial domain of the problem into smaller components to approximate different expressions in (2.38), the spatial precision can be enhanced by refining the grid *i.e.* dividing the spatial domain into more component cells and thus introducing a finer mesh for making the required approximations. Of course, one might argue that using smaller time steps would increase the accuracy of the numerical approximation of the solution as well; this is a justified argument, however, since the numerical method used for the current problem is explicit, which means it already requires comparatively small time steps for stability (associated with high computational cost), we would not concern ourselves with showing convergence in that sense.

The inherent stochasticity in our model, introduced by the special treatment of random attachments, renders the solution attained by a particular individual simulation of (2.38) to be a single sample from all the possible solutions. In other words the solution and especially two of its components that are important from a biological point of view, namely the biomass density and the carbon concentration, are fairly sensitive to the rate, spatial position and spatial spread of the randomly attached cells. The rate of these random attachments can be partly controlled by manipulating the constant coefficient, $\psi$, for a specific grid size. However, as the grid is refined, the rate of attachments increases as well since our simulation code treats each grid cell as a viable attachment point and hence the larger the grid size, the more the instances of cell attachments. This issue could be addressed by dynamically adjusting $\psi$ so that the same rate of attachment is roughly maintained throughout the refinement procedure. But since we will study the effect of this parameter later in the sensitivity analysis section, we do not apply this solution here. Instead of changing the value of $\psi$ for each grid size, we attempt to maintain the spatial properties of the random attachments constant throughout the refinement procedure so that on average the rate of attachments is relatively unchanged. The spatial position of random attachments has to be kept random since it is the main goal of studying the stochastic cell attachment. As for the spatial size of a single attachment instance, it would be the same size as a single grid cell, therefore if we keep the same manner of attachment mechanism as discussed before, for more refined grids the spatial size of a single attachment would shrink with the same factor. In order to have control over this aspect of the random attachments, we have introduced some modifications in the parallel
section of the code: instead of generating a normally distributed random number per grid cell, the same random number is used to populate a group of contiguous cells (depending on the assumed spatial size of attachments). This way we are able to keep the spatial size of the attachments the same for all the grid sizes. Figure 4.8 and Figure 4.9 are individual samples of many simulations with the same $\psi$ values and spatial length of cell attachments.

Since the value of the biomass density and the carbon concentration inside a particular grid cell takes a specific trajectory through time – which is influenced by a few external factors such as the simultaneous state of the adjacent cells – a more efficient and less complex approach to study the state variables of interest would be studying the spatial integration of these variables over the whole domain at a number of time points. This way we can reduce the degrees of freedom in the solution components and thus the convergence results become more amenable to representation in both numerical and graphical modes. The following formulae are used to spatially integrate the biomass density and carbon concentration

$$M_{\text{TOT}} := \int_{\Omega} M(t, x) dx = \Delta x \sum_{p=1}^{NM} M_p,$$

and

$$C_{\text{TOT}} := \int_{\Omega} C(t, x) dx = \Delta x \sum_{p=1}^{NM} C_p,$$

where $\Omega$ is the spatial domain and $NM$ is the grid size ($M = 1$ in 1D). Here, $p$ is the single index that designates each grid cell using index mapping $\pi$ defined in Section 2.1.3. These formulae are applicable to both 1D and 2D settings where $\Delta x$ is the length of a grid cell in 1D and the length of a side of a grid cell in 2D (see Section 2.1.3 for the precise definition of $\Delta x$).

As an example for this treatment of the state variables, we have provided the spatially integrated biomass density and carbon concentration in Figure 4.11 and Figure 4.12 for two different attachment sizes. A close examination of these curves reveals the stochastic nature of the studied variables. Although one can distinguish a fair level of convergence in Figure 4.11, the same does not hold true after examining Figure 4.12.
$\psi = 0.2$, $sd = 16$

Figure 4.11: The time evolution of the total biomass (left) and carbon (right) for several grid sizes. The spatial size of each random cell attachment is $\frac{1}{16}$ of the domain length.

$\psi = 0.1$, $sd = 16$

Figure 4.12: The time evolution of the total biomass (left) and carbon (right) for several grid sizes. The spatial size of each random cell attachment is $\frac{1}{128}$ of the domain length.

This difference between the simulations depicted in these figures can be generally ascribed to the spatial size of the attachments and the number of grid cells. For spatially larger random attachments a greater proportion of the domain is covered by initial colonies and as a result the domain becomes less available to further random attachments. In contrast, the spatially smaller random attachments cover a lesser proportion of the domain and even after several instances of nascent colonies' emergence, there is still room for further random attachments. The simulation time for the case of smaller random attachments becomes longer as a result
(the time it takes for both the biomass and carbon to vanish), as long as the rate of random attachments is kept sufficiently small. The longer simulation time requires more time steps, which means more sets of random numbers are generated. This can potentially result in slightly positionally different random attachments, especially for finer grids for which the time steps are very small.

In order to establish whether these variables show any sign of convergence in a statistical sense, multiple instances of the same data type should be generated so that some statistics such as the sample expected values (mean) and sample standard deviations can be calculated.

Figure 4.13 and Figure 4.14 show the biomass and carbon mean for a hundred realizations of each grid size.

![Figure 4.13: The time evolution of the total biomass (left) and carbon (right) for several grid sizes averaged over a hundred realizations. The spatial size of each random cell attachment is $\frac{1}{16}$ of the domain length. Here, $\psi = 0.2$.](image)

The convergence of the state variables is notably more distinguishable when the mean values are considered rather than data from a single simulation. We have calculated the sample standard deviations and the standard error of the sample mean for all the grid sizes as well. We have used the following formulae to calculate these statistics

\[
\text{sample mean} := E[X] = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad (4.3)
\]

\[
\text{sample standard deviation} := SD[X] = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - E[X])^2}, \quad (4.4)
\]
standard error of the sample mean := $SEM[X] = \frac{SD[X]}{\sqrt{n}}$, \hspace{1cm} (4.5)

where $X$ can be any of the spatially integrated state variables and $n$ is the number of samples (realizations). Using the values obtained for the standard error of the sample means, we have calculated the distribution of data around the sample means. These are displayed in Figure 4.15 and Figure 4.16.
Figure 4.15: Spatially integrated biomass density (blue) and carbon concentration (red) shown for several grid sizes and averaged over a hundred simulations with spatial size of attachments $\frac{1}{16}$ of the domain length. A vertical line segment of the light colored areas is an error bar with length twice the magnitude of the standard error. The insets are magnifications of the areas around the peak of the biomass curves, which provide a more resolved view. The rest of parameters are the same as before.
Figure 4.16: Spatially integrated biomass density (blue) and carbon concentration (red) shown for several grid sizes and averaged over a hundred simulations with spatial size of attachments $\frac{1}{128}$ of the domain length. A vertical line segment of the light colored areas is an error bar with length twice the magnitude of the standard error. The insets are magnifications of the areas around the peak of the biomass curves, which provide a more resolved view. The rest of parameters are the same as before.
Figure 4.15 and Figure 4.16 include insets that show a magnified section of the curves. By examining these insets, one notices that the length of error bars (vertical line segments that collectively constitute the light colored areas around the curves) shortens progressively as a function of refinement. This effect is more pronounced for the biomass curves in comparison with the substrate curves. In order to support the above claim, we have plotted the values of the standard error of the sample mean for each of the grid sizes against time (see Figure 4.17 and Figure 4.18).

Figure 4.17: The biomass SEM (left) and carbon SEM (right) for several grid sizes obtained from a hundred realizations. The spatial size of each random cell attachment is $\frac{1}{16}$ of the domain length. Here, $\psi = 0.2$.

Figure 4.18: The biomass SEM (left) and carbon SEM (right) for several grid sizes obtained from a hundred realizations. The spatial size of each random cell attachment is $\frac{1}{128}$ of the domain length. Here, $\psi = 0.1$. 
By examining the above figures, one can notice that as finer grids are used for numerical calculations, the standard errors of the sample mean for both of the state variables show convergence to certain profiles.

Up to this point, we have solely used graphical representations to demonstrate the grid-refinement convergence of the biomass density and the carbon concentration for two different spatial sizes of random cell attachments. Next we provide a quantitative approach for showing these results. This involves calculating the 2-norm and 1-norm of the statistics we have incorporated so far. We deploy the norm calculations in three different ways. First, we calculate the 2-norm of the difference between the mean values of the spatially integrated solution components, $E[M_{TOT}]$ and $E[C_{TOT}]$, of two consecutive grid sizes (see Figure 4.19). Second, we implement the previous operation a bit differently such that instead of finding the difference between consecutive grid sizes, it is found by subtracting the data for the finest grid size from the remaining grid sizes (see Figure 4.20). Third, we calculate the 1-norm of the standard deviations associated with each value of the spatially integrated solution component (see Figure 4.21). For the sake of brevity in notation use, we have omitted the subscript $TOT$ in the following figures.

The results of the first two treatments where 2-norm operations are used, manifest grid convergence for both the mean biomass density and the mean carbon concentration; we can spot a definite reduction in the values of the calculated 2-norms of the differences in the means for both cases of random attachment sizes. A similar reduction in the value of the calculated 1-norm of the standard deviations can be observed for the biomass density; however, the carbon concentration does not show the same behavior: we speculate that this can be attributed to the slower rate of convergence for carbon solutions as observed (and privately communicated) by Eberl et al. in simulations of (2.38) [12]. It should also be noted that the values of the biomass 1-norms begin to level off as the grid size increases in both cases of random attachment sizes.
\( \psi = 0.2, \) random attachment size = \( \frac{1}{16} L, k = 4, ..., 12 \)

\( \psi = 0.1, \) random attachment size = \( \frac{1}{128} L, k = 7, ..., 13 \)

Figure 4.19: 2-norm of the difference between the values of the spatially integrated solution components of two consecutive grid sizes plotted against the base-2 logarithm of the difference between grid sizes. The spatial size of random attachments for the top figures is eight times as large as the bottom ones. Number of samples used: 100.
ψ = 0.2, random attachment size = $\frac{1}{16}L$, $k = 4, \ldots, 11$

ψ = 0.1, random attachment size = $\frac{1}{128}L$, $k = 7, \ldots, 12$

Figure 4.20: 2-norm of the difference between the values of the spatially integrated solution components of each grid size and the finest grid plotted against the base-2 logarithm of the grid sizes. The spatial size of random attachments for the top figures is eight times as large as the bottom ones. Number of samples used: 100.
$\psi = 0.2$, random attachment size $= \frac{1}{16} L$, $k = 4, \ldots, 12$

$\psi = 0.1$, random attachment size $= \frac{1}{128} L$, $k = 7, \ldots, 13$

Figure 4.21: 1-norm of the standard deviations of the spatially integrated solution components plotted against the base-2 logarithm of the grid sizes. The spatial size of random attachments for the top figures is eight times as large as the bottom ones. Number of samples used: 100.
4.2.2 Refinement and Stochasticity

As we mentioned earlier, the grid refinement has a direct effect on the rate of random cell attachments; due to the spatial discretization of the model, each grid cell has its own AF, hence in our computational realization of the model the spatial size of the smallest random attachment is bounded from below by the size of a single grid cell; as we introduce more grid cells as a function of refinement, the number of spots available for potential attachments by free cells in a single time step grows as well. This can also be understood by considering what is taking place inside each time step: for each time iteration we require a new random number for each grid cell in order to be able to simulate the corresponding Wiener increment; since there are more grid cells inside a finer grid, there is a greater chance for having a larger number of corresponding AFs to satisfy the attachment conditions.

By increasing the instances of stochastic attachments as a result of refining the grid, the spatial separation of initial colonies dwindles. This can lead to a more rapid coalescence of these nascent colonies. Therefore, the formation of a relatively large community that covers most of the domain in a short amount of time is more probable for the finest grids (see Figure 4.10).

Considering the aforementioned effects of refinement on the frequency of random attachments, the next step would be studying how refinement affects the inherent stochasticity of the model. We especially want to investigate this effect by maintaining the biomass amount of randomly attached cells approximately the same over equal periods of simulation times: since larger grids are associated with smaller amounts of biomass per grid cell and smaller time steps, we expect that the amount of added biomass over a particular interval of time stays the same on average for all grid sizes. This way we can isolate the stochastic effects of the spatial distribution of the random attachments and initial colonies without any concerns over drastic changes in their overall biomass values. Due to stochastic nature of these experiments, a statistical approach for analyzing the simulation results is required. Again, we provide the results for the spatially integrated state variables.

Figure 4.22 displays the spatially integrated biomass density and carbon concentration at different time points that are averaged over a hundred realizations. The grid refinement and the concomitant intensification of the random attachment rate have discernible impact on the shape of the depicted curves. The mean total biomass curve becomes narrower, its peak grows higher and the peak value is attained earlier for larger grid sizes. This can be explained by our observations from Figure 4.10, i.e. the more the instances of random attachments with close proximity, the faster the consolidation of smaller colonies and the
Figure 4.22: The time evolution of the total biomass (left) and carbon (right) for several grid sizes averaged over a hundred realizations. The spatial size of each random cell attachment is $L/N$. Here, $\psi = 0.2$.

narrower in time and more homogeneous in space the peak activity of the domain-covering community. As for the mean carbon curves, we observe that the grid refinement increases the rate of substrate consumption (carbon vanishes faster in time), which is in accordance with our observations of the biomass curves.

Using (4.3), (4.4) and (4.5) we have calculated the standard error of the sample mean for both state variables. Figure 4.23 provides the graphical representation of the same variables as before with the addition of data distribution (light colored areas) around the sample means. The insets are magnifications of smaller sections of these plots, which provide a better view of the error in data. The most conspicuous feature in these insets is the gradual shortening in the length of the error bars (vertical line segments that collectively constitute the light colored areas around the curves) as the number of grid cells is increased.
Figure 4.23: Spatially integrated biomass density (blue) and carbon concentration (red) shown for several grid sizes and averaged over a hundred simulations. A vertical line segment of the light colored areas is an error bar with length twice the magnitude of the standard error. The insets are magnifications of the areas around the peak of the biomass curves, which provide a more resolved view. The rest of parameters are the same as before.
If we plot the standard error of the sample mean against time for each grid size, we obtain the graphs shown in Figure 4.24.

A visual examination of the above graphs reveals a systematic decline in the value of the standard error of the sample mean at all time points as the grid size gets larger. This behavior can be interpreted as a lowering in the level of stochasticity that affects the overall trajectory of the state variables. In other words, when the frequency of stochastic attachments is relatively high such that in a short amount of simulation time most of the domain is occupied by small colonies with close proximity to one another, the time evolution of the spatially integrated state variables does not change significantly from one simulation to the next as long as all the other parameters are kept the same. Therefore, a reasonable conjecture based on these observations would be the following: when the biofilm formation is initiated by numerous colonies, which are themselves established by a large number of random cell attachments that have occupied the available substratum in a somewhat homogeneous manner, the time evolution of the spatially integrated state variables predicted by the stochastic model can also be replicated by the deterministic model.

4.3 Parameter Sensitivity Analysis

Here we investigate the dependence of the model behavior on some of the model parameters, specifically the ones that involve the stochastic attachment of cells. Unlike the approach
adopted in the previous sections, we have implemented the parameter sensitivity analysis in a 2D setting since the spatial grid used for these simulations has been set to a fixed size and as a consequence we have opted for a grid size that is not computationally too expensive for numerical calculations. The chosen grid size is $256 \times 256$, which provides a sufficiently refined grid with relatively cheap computational cost.

In order to examine the model sensitivity to its parameters in the present setting, we have focused on the quantitative variations the state variables manifest upon slight changes in the value of a single parameter. As before, instead of considering the values of state variables as functions of position and time, the spatially integrated values have been used. Since the model is inherently stochastic, we have run the simulations a number of times and have applied a statistical treatment to the obtained results (as we did in the previous section).

The parameters that are the focus of sensitivity analysis are the characterizing parameters of the impulse function namely, $a$, $b$ and $\sigma$, and the constant coefficient $\psi$ that appears in (2.35). We conjecture that these parameters possess the highest degree of influence on the stochastic behavior of the model. The rest of the parameters are either deemed not as influential or not related to the stochastic cell attachment; the values of all parameters that are not subject to sensitivity analysis are given in Table 4.1.

Two of the characterizing parameters of the impulse function, i.e. $a$ and $b$, can be studied in different ways. Here we have decided to keep a constant length for the interval that characterizes the effective domain of the impulse function. For sufficiently small values of $\sigma$ this length equals $b - a$. In order to study the effect of changing the values of $a$ and $b$, we have incremented these parameters by certain values with respect to the initial data assumed for the AF while the interval length, $b - a$, has been maintained the same.

In the following we introduce a simulation scenario for conducting sensitivity analysis. The accompanying figures include the changes of model behavior due to input parameters as well as the magnitude of stochastic effects using the usual statistics.

For all sampled simulations, the substratum or spatial domain has been set to have zero biomass attached to it initially. As simulation time progresses, random cell attachments occur, and subsequently establish small colonies, which then grow, consume the substrate and eventually merge to form larger communities (Figure 4.1 and Figure 4.2 provide visual representations for this scenario).

Parameters $a$, $b$, $\sigma$ and $\psi$ have been incremented by arbitrary values (chosen by examining many test simulations and considering the time constraints and computational costs) in
order to study their effect on the time evolution of the spatially integrated state variables. It should be noted that we have studied the effect of $a$ and $b$ in tandem as explained above. To incorporate the stochastic fluctuations in our results, a hundred simulations for each value of the studied parameters have been executed. Formulae (4.3) and (4.4) have been used to calculate the means and standard deviations at different time points. We have used the standard deviation values to show the level of stochasticity in the spatially integrated $M$ and $C$. This is represented by vertical bars in the following figures, which are formed by adding and subtracting one standard deviation at a number of points on the curves.

The effects of incrementing $\psi$ are shown in Figure 4.25. As we elaborated in Section 4.1.1, the higher values of this parameter are associated with more instances of stochastic cell attachments that also begin to occur earlier in time. This model behavior is corroborated by the results shown here as well. We can clearly see that both biomass and carbon curves attain their maximum and minimum values, respectively, earlier in time as higher values of $\psi$ are deployed. By invoking our previous interpretations, this outcome is definitely in accordance with the expected behavior: the greater $\psi$ values result in more random cell attachments and as more initial colonies form, the substratum is covered by more distributed amounts of biomass, which speeds up both the growth phase and the consumption of the carbon substrate. As a result, the biomass curves become narrower and attain higher peaks, and the carbon curves become steeper during the consumption phase.

![Figure 4.25: The mean values of the spatially integrated biomass density (left) and carbon concentration (right) plotted against time for several values of $\psi$. Here $a = 0.99$, $b = 1.00$ and $\sigma = 0.00001$ for all the one hundred sampled simulations. The vertical bars represent the mean ± one standard deviation.](image)

Another noticeable feature is the gradual contraction of the vertical bars as the value of $\psi$ increases.
ψ is increased. This means that the numerical value of the standard deviation is decreasing. This situation is similar to the one discussed in Section 4.2.2. As the number of colonies grows due to higher rates of random attachments the uncertainty in $M_{TOT}$ and $C_{TOT}$ curves is reduced. In other words, the stochastic fluctuations from one sample to the next become relatively less noticeable; that is, when most of the substratum is covered by biomass in a short period of time, the growth and death phases have fewer possible number of outcomes probabilistically, and so is true for the possible outcomes of substrate consumption, i.e. fewer outcomes; as a consequence the dispersion of the gathered data becomes much narrower around the means.

Figure 4.26 shows the effects that different values of $a$ and $b$ cause in the time evolution of the total biomass and total carbon. One can easily verify that $b - a = 0.05$ for all the values of these parameters. As we reduce the values of $a$ and $b$ while $b - a$ is kept constant, the maximum and minimum values of the biomass and carbon curves, respectively, are attained earlier in time. This behavior is similar to the one manifested by higher values of ψ, which induced more frequent random attachments. Since the value of ψ is the same for all the sampled simulations in this case, the reason that more attachments occur for smaller values of $a$ and $b$ may be attributed to the higher likelihood of the event where the value of the AF falls within the impulse interval as the interval is shifted towards zero (the initial value of the AF over the whole domain).

Figure 4.26: The mean values of the spatially integrated biomass density (left) and carbon concentration (right) plotted against time for several values of $a$ and $b$. Here $ψ = 0.1$, $b - a = 0.05$ and $σ = 0.00001$ for all the one hundred sampled simulations. The vertical bars represent the mean ± one standard deviation.

We also observe the same gradual decrease of the numerical value of the standard
deviation, which is represented by shorter vertical bars in the graphs. The reason for it is the same as the one discussed for $\psi$ i.e. more instances of random attachments cause the uncertainty to decline.

The situation is a bit different in the case of incrementing the value of $\sigma$ (at least for the values used here). As Figure 4.27 shows, the curves plotted for different values of $\sigma$ overlap one another for the most part and slight discrepancies only emerge for the last couple values of $\sigma$, namely 0.01 and 0.001. We have provided a zoomed-in portion of the graphs for more clarification. Obviously the numerical value of this parameter does not impact the frequency of stochastic attachments as much as the other three parameters do. The reason for the fairly low sensitivity observed here can be ascribed to the shape of the impulse function and its dependence on $\sigma$. Referring back to Figure 2.2, we notice that for $\sigma$ values close to the ones used here the impulse function maintains a rectangular shape between the values of $a$ and $b$ where the sides meet the horizontal axis very close to the exact values of $a$ and $b$. Hence, the cut-offs for effective $\phi$ values are sharp and very close to the numerical values of $a$ and $b$. For relatively greater values of $\sigma$ the rectangular shape gives way to a mound-like shape with variable width, which does not equal the length of $[a, b]$ everywhere. Therefore, as long as $\sigma$ assumes values in the range that maintains somewhat sharp cut-offs for $\phi$, the frequency of stochastic cell attachments are most sensitive to the values of $a$, $b$ and $\psi$.

![Figure 4.27](image_url)

Figure 4.27: The mean values of the spatially integrated biomass density (left) and carbon concentration (right) plotted against time for several values of $\sigma$. Here $a = 0.99$, $b = 1.00$ and $\psi = 0.1$ for all the one hundred sampled simulations. The vertical bars represent the mean ± one standard deviation. The insets provide a zoomed-in portion of the graphs for clarification.

As for the stochastic fluctuations experienced by the state variables $M_{TOT}$ and $C_{TOT}$, one
notices no dramatic changes in the length of the vertical bars. This observation is validated by the fact that the tested \( \sigma \) values have little to no effect on the number and frequency of random attachments as discussed above. Thus, the standard deviation values obtained for the lowest \( \sigma \) value remain somewhat unchanged through all the other values.
Chapter 5

Discussion

Through performing many numerical simulations, we have shown that the presented approach for modeling the stochastic attachment of free bacterial cells to the substratum is a viable mathematical framework. We especially focused our modeling strategies to the case of cellulolytic biofilms for which a deterministic PDE-ODE system has been proposed to model the biomass growth, the inverse colony formations and the nutrient uptake [12].

The particular nature of the random cell attachments and the fact that there are not sufficient experimental data in the literature to be used for proposing some possible underlying mechanisms that drive the cellular attachments, especially in the case of cellulolytic biofilms, make the deployment of a stochastic approach for modeling the attachment process adequate for now. However, future experiments and further analysis of their results may eventually provide a path towards a model that incorporates the cellular attachments in a fully deterministic way.

As it turns out the mere addition of the SDE framework to the model does not lend itself readily and directly to including the random cell attachments. Primarily, the inherent properties of the Wiener process that drives the stochastic part of the Itô SDEs, make the simple conversion of the biomass PDE into an SPDE problematic: the possibility of negative solutions and the random magnitude of the stochastic part render this approach unsuitable for constructing random attachments with certain characteristics. This is the main reason as to why we have decoupled the stochasticity from the main model equations and built it as a separate variable, namely $\phi$.

The utilization of the theoretically well-established formalism of the stochastic differential equations as the governing force of stochastic cell attachments is a step forward from the ad-hoc method used in [12]. Although this can be viewed as one of the most significant
accomplishments of the present work, one has to admit that the underlying stochastic variable, namely the attachment factor denoted by $\phi$, which produces the required stochasticity in the system, is not physically on par with the other two state variables, the biomass density $M$ and the carbon concentration $C$. It appears as a mathematical ploy, for lack of a better word, that makes the incorporation of the SDE formalism possible in the present context. But since our main purpose for this project has been utilizing an approach that is bolstered by a rigorous theoretical foundation, the introduced SDE formulation can be viewed as the first step towards modeling the stochastic attachment phenomenon. Here we acknowledge that there is room for improving this particular approach.

After spatial discretization of the governing PDE-ODE system, the biomass ODEs are coupled to the Itô SDEs by an impulse function. This function is the principal mechanism that causes the attachment of a certain amount of biomass to a position on the substratum depending on whether sufficient space and nutrient are available locally. The AF, represented by $\phi$, performs as the input signal to the impulse function and for certain values of it an impulse is switched on, which one can control by adjusting the defining parameters, and the corresponding location on the substratum is inoculated by a tiny amount of biomass $i.e.$ a cell attachment occurs. As long as the stochastic process that governs the AF is random enough, one has no a priori knowledge as to when and where the attachment events may take place. We conjecture that the key element for simulating these discrete attachment events with their specific properties is the impulse function or rather the impulsive dynamics; even if all the factors influencing the attachment of bacteria were known and modeled using differential equations, the need for an impulsive mechanism that accepts these factors as its input would still be essential since the cellular attachments are discrete events in both space and time and have an impulsive nature to them. One of the other benefits of deploying an impulse function that has a nonnegative range is the vanishing of the negative biomass solutions. This was a major point of concern especially with the direct addition of noise to the biomass ODEs.

The underlying assumption wherever we have plotted the temporal behavior of the biomass density and the carbon concentration is the existence and uniqueness of these solution components. We have based our assumption mainly on the proof of existence and uniqueness for the diffusion-reaction equations with diffusive substrate [16] as well as the existence, uniqueness, non-negativity and boundedness of the solutions of the regularized system of ODEs that result from numerical treatment of the main equations [21]. In Proposition 1, we have shown the non-negativity and boundedness of the solution components
\( M^* \) and \( C^* \) of the regularized problem by invoking suitable theorems from \[9\]. Although the boundedness of the biomass density and carbon concentration in the regularized version of our model is guaranteed by the proof of Proposition 1, further analysis is still needed for establishing the existence and uniqueness of solutions\(^1\) and the effects of the impulsive component of the model on the mathematical properties of the whole system. Of course these investigations were beyond the scope of this thesis and would be of great importance for elucidating the mathematical underpinnings of these types of models much further.

The numerical simulations of our model for stochastic cell attachments have demonstrated a number of attributes. For instance, we have shown that in a two dimensional setting, the proposed approach for emulating the random attachment of free cells to the surface of the substratum produces graphical results that exhibit similar characteristics to the cellulolytic biofilm formations shown in Figures 1.3 and 1.4 \[57\]. Furthermore, our 2D simulations may be viewed as computer experiments where different scenarios with various initial states can be replicated and consequently an improved understanding of the studied biological system could be acquired. In terms of numerical analysis, the grid convergence of the averaged spatially integrated biomass density and carbon concentration has been shown for the one dimensional setting. We assume, the investigation into the 2D convergence follows the same path as the 1D one, however, the sheer amount of computational cost incurred by repeating the simulations for each grid size with different random numbers, several times, which is needed in order to be able to analyze the results statistically, has deterred us from conducting such studies. These have the potential to be the topic of computational analysis in subsequent work.

The number and frequency of stochastic cell attachments is most sensitive to the constant coefficient in the AF SDE, \( \psi \), and the shift parameters of the impulse function, \( a \) and \( b \). When the numerical value of \( \psi \) is relatively high or \( a \) and \( b \) shift the effective domain of the impulse numerically towards the initial data for \( \phi \), the observed overall behavior is more instances of cell attachments close to each other both spatially and temporally and the emergence of small colonies earlier in time. As \( \psi \) is set to smaller values or the effective interval is shifted away from the initial data for the AF, the attachments become more sparse in space and time and the emergence of the first colonies begins later in time. Therefore, we can conclude that the developed stochastic approach manifests an intertwined spatiotemporal characteristic meaning that the time component of attachments and their spatial component

\(^1\)It has been suggested to us that the existence and uniqueness of solutions of the regularized system of SDEs (2.40) should follow from Theorem 5.2.1 in \[47\].
cannot be controlled by the above parameters independent of one another at least within the current framework and up to our knowledge to this point\textsuperscript{2}.

As a direct consequence of the model being inherently stochastic, some of the simulation results and drawn conclusions have been treated statistically. The number of sampled simulations for these statistical treatments have been determined such that the level of stochasticity would not change dramatically as more samples were included in calculating the desired statistics. For determining this number, we have used the 1-norm of the standard deviation as a measure, which can be calculated from the data for both 1D and 2D simulations.

In the case of 1D simulations, we have both biomass data and carbon data for a number of grid sizes that have been used for the results in Section 4.2.1. The 1-norm is calculated easily by summing over standard deviations from several time points. Figure 5.1 shows the results of these calculations for one of the grid sizes. we have chosen the x-axis to represent the number of samples. As the sample size is increased, the 1-norm of the standard deviation

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5.1.png}
\caption{The 1-norm of standard deviation calculated using standard deviations at several time points for a 1D setting. The x-axis represents the number of samples. The grid size is $2^{14} = 4096$. The biomass curve is on the left and the carbon curve is on the right. Note that the numerical data from the spatially integrated state variables are used here. The simulation parameters are given in Table 4.1 and specifically here $\psi = 0.2$, $a = 0.99$, $b = 1.00$ and $\sigma = 0.00001$.}
\end{figure}

for the biomass begins to show less variation above seventy samples. Therefore, for keeping the computational costs manageable, we have used a maximum of one hundred samples for

\textsuperscript{2}We have shown a possible mechanism for varying the spatial rate of random attachments independent of their temporal component by manipulating the initial data for the AF, but since we have not examined this thoroughly from a numerical analysis point of view, we refrain from further assertions in this regard.
calculating the desired statistics. The same type of behavior is somewhat evident from the carbon curves as well. This result holds true for other grid sizes, which we have not displayed here.

The 2D case, is handled similarly. The biomass and carbon data have been acquired from sensitivity analysis simulations as described in Section 4.3. Figure 5.2 shows the results for calculating the 1-norm of standard deviations for different sample sizes. Here, like the 1D case, the x-axis represents the number of samples. As we can clearly see, both biomass and carbon curves begin to show signs of leveling off after the number of samples exceeds eighty. Hence, for keeping the computational costs as low as possible, we have deemed a hundred samples a suitable sample size for our experiments in 2D simulations.

![Figure 5.2: The 1-norm of standard deviation calculated using standard deviations at several time points for a 2D setting. The x-axis represents the number of samples. The grid size is 256×256. The biomass curve is on the left and the carbon curve is on the right. Note that the numerical data from the spatially integrated state variables are used here. The simulation parameters are given in Table 4.1 and specifically here $\psi = 0.08$, $a = 0.99$, $b = 1.00$ and $\sigma = 0.00001$.](image-url)

One of the most challenging tasks in this project has been the computational realization of the numerical simulations. This has manifested itself at different levels e.g. the inherent instabilities of the explicit numerical methods and the process of random number generation.

The algorithms we have utilized for random number generation are textbook methods with well-known properties. However, using these algorithms in the current context is not warranted to be flawless since due to relatively small time steps required for the stability of the integration methods, a huge number of random numbers must be generated to fill up the spatial grid. For one thing the pseudo random number generating algorithms may not
possess the expected statistical properties for this level of demand such that running into periodic strings of random numbers under these conditions would be fairly likely. The other flaw could be arising from the computer code itself as we have developed all the code used for this work in-house and the issues with efficiency of our code for these algorithms can have major impacts on the runtime speed of the simulations. For these reasons, we recommend the usage of math libraries and pseudo random number generator codes that are tested and proven to be efficient especially for parallel computing purposes.

The issue with the time step sizes had to be dealt with in a bit unorthodox manner so that simulations maintain a meaningful temporal progression of values for both the biomass density and carbon concentration. The Euler method is classically defined for fixed time steps. But in our computational approach, we could not retain the accepted definition and had to change the time step sizes based on the highest value of the biomass density over the whole grid within each time step. This specific approach was adopted for two reasons: first, for the stochastic changes in the amount of biomass that a grid cell might experience and second, for increasing the runtime speed. The stochastic attachments, especially in grid cells that have biomass densities close to one, can easily throw the system into instability. This is mainly because the upper bound (3.13) on the step size is inversely proportional to the local diffusion, which for values of the biomass density close to one is large; thus, the calculations for the current values of the state variables are very sensitive to the estimate for the time step size. One could argue that the criterion (3.13) could be used only once to determine an estimate of the smallest time step for a value of biomass density that is deliberately chosen very close to one and then treat the obtained estimate as the fixed step size for the Euler method. However, this can easily work for one simulation and fail for the next sampled simulation because of the biomass density fluctuations. Moreover, running the whole simulation by relatively small time steps turns out to be very expensive computationally, especially for relatively large grid sizes. As a result of these considerations, we have implemented the dynamic calculation of the time step sizes the way we laid out at the end of Chapter 3. To sum it up, our computational approach restarts the Euler method in every time step.
Chapter 6

Conclusion

In the following we point out a couple of most salient observations we have made conducting this research and finally, we provide some of our preliminary ideas that a future work could utilize for exploring new avenues within the proposed modeling framework.

6.1 Lessons Learned

- Discrete random cell attachment can be included in a deterministic continuous biofilm model by introducing an auxiliary stochastic process that generates attachment impulses. This formulation leads to a stochastic model that preserves non-negativity as opposed to the direct addition of white noise. In contrast to an earlier more direct approach to include (nonlinear) stochastic impulses, this formulation leads to an Itô stochastic differential equation for which the established stochastic numerical methods are known to reproduce the stochasticity of the differential equation correctly.

- Numerically, the explicit time integration schemes deployed for this work turned out to be highly expensive. We acknowledge that more advanced deterministic and stochastic time integration schemes should be explored.

6.2 Future Work

One possible path for future work on our proposed stochastic model could be attempting to apply it to a biofilm structure that in addition to the two dimensional growth over the substratum, as is the case for cellulolytic biofilms in this thesis, also undergoes growth in
the third spatial dimension \textit{e.g.} into the aqueous phase. By invoking similar treatments performed in Chapter 2 for spatial discretization, the 3D domain can be discretized, in the same manner, by a spatial grid where the grid cells are going to be rectangular prisms. We speculate that due to the more complex structural mechanics of biofilm formation in 3D, the numerical simulation should be divided into a few parts where each part executes a set of instructions depending on the specific conditions that have to be taken into consideration. These conditions are imposed by factors such as whether there is a physical surface onto which a free cell can randomly attach and whether the cell can physically access that position. For instance, in a 3D setting, each grid cell has six neighbors that share a side with it (five neighbors for cells at the faces of domain boundaries, four for cells at the edges of domain boundaries and three for cells at the corners or the vertices of domain boundaries). When there is some biomass in any of these neighbors or when the grid cell is adjacent to the substratum, some structural base for free cells to attach on exists; however, if all the neighbors are occupied by some amount of biomass then the physical access to the grid cell is presumably cut off from the outside world. Depending on what the circumstances are, different sets of equations have to be called up for numerical calculations; for example, when there is no physical access to a grid cell for bacterial attachment, its temporal evolution ought to be solely governed by the deterministic equations, which should be also the case for a grid cell that is surrounded by empty neighbors and away from the substratum.

We also speculate that the nutrient delivery in a 3D structure should probably take place via other mechanisms as well. We know that most of the grid cells are not adjacent to the substratum and hence, there can be no direct nutrient absorption by any biomass inside them. Furthermore, those grid cells that contain biomass but are shielded by other biomass-containing neighbors would have no access to nutrients at all. Therefore, for these situations it seems to be necessary to consider diffusing substrates, for the growth and overall sustainability of the biofilm structure, that are transported by or dissolved in the surrounding fluid.
Bibliography


