



Research article

A mathematical model of discrete attachment to a cellulolytic biofilm using random DEs

Jack M. Hughes¹, Hermann J. Eberl^{2,*} and Stefanie Sonner³

¹ Department of Mathematics, University of British Columbia, 1984 Mathematics Rd., Vancouver, BC V6T 1Z2, Canada

² Department of Mathematics and Statistics, University of Guelph, 50 Stone Rd E., Guelph, ON N1G 2W1, Canada

³ Radboud University, IMAPP – Mathematics, Postbus 9010, 6500 GL Nijmegen, The Netherlands

* **Correspondence:** Email: heberl@uoguelph.ca; Tel: +15198244120 ext. 52622.

Abstract: We propose a new mathematical framework for the addition of stochastic attachment to biofilm models, via the use of random ordinary differential equations. We focus our approach on a spatially explicit model of cellulolytic biofilm growth and formation that comprises a PDE-ODE coupled system to describe the biomass and carbon respectively. The model equations are discretized in space using a standard finite volume method. We introduce discrete attachment events into the discretized model via an impulse function with a standard stochastic process as input. We solve our model with an implicit ODE solver. We provide basic simulations to investigate the qualitative features of our model. We then perform a grid refinement study to investigate the spatial convergence of our model. We investigate model behaviour while varying key attachment parameters. Lastly, we use our attachment model to provide evidence for a stable travelling wave solution to the original PDE-ODE coupled system.

Keywords: attachment; cellulolytic biofilm; cellulolytic biofilm; mathematical model; nonlinear diffusion; numerical simulation; random ordinary differential equation

1. Introduction

Bacterial biofilms are microbial depositions on immersed solid abiotic or biotic surfaces. Once attached, they form a self-secreted matrix of extracellular polymeric substance (EPS) that provides structural stability and an adhesion mechanism to the surface. The EPS is assumed to be mostly composed of polysaccharides, but it may contain nucleic acids and proteins. The bacterial cells become sessile and are protected from anti-microbials and mechanical stress due to the EPS [1–3].

The surface to which the biofilm attaches is called the substratum. Some substrata are impenetrable and inert such as rocks or plastic. However, we will be considering substrata that are reactive and can be degraded. These degradable substrata can play a vital role in the biofilm development cycle as they can provide the nutrients needed for biofilm growth and proliferation, which is the main focus of the cellulolytic biofilms we will focus on.

Biofilms require nutrients to grow, which we call substrates. Substrates are either supplied by the substratum (this is the case for cellulolytic biofilms) or by the surrounding environment through diffusion or advection. Depending on the availability, rate of delivery, and local concentrations of substrates, biomass growth may be slow and limited in specific locations, while in other locations, the biomass may be fast-growing and abundant. If we assume the availability of substrates is not a growth-limiting factor, then the metabolic growth rate of the biofilm will determine the generation of biological by-products, and the uptake of substrates [4].

A biofilm community typically consists of many separated sub-colonies, which may merge and form into larger colonies. Macroscopically, biofilm communities are homogeneous, meaning they look like a thin film covering the substratum. However, structural differences become apparent at the mesoscopic level ($10\mu\text{m}$ – 1mm in this case). At these scales, the heterogeneity of a biofilm community manifests itself. Factors such as specific growth rate, maximum cell density, and substrate concentration determine the level of heterogeneity of the biofilm community.

The formation of biofilms onto the substratum typically occurs via the physical attachment of free-floating bacteria in the aqueous phase surrounding the substratum by forming a firm physiochemical bond. One possible source of free-floating bacteria is from biomass detachment as a result of shear forces exerted on the biofilm from a fluid [5]. The possibility of adhesion of bacterial cells to a certain position on the substratum can be influenced by many factors, including [6]:

- Bacterial characteristics: surface charge, appendages, hydrophobicity, etc.,
- Environment: temperature, bacterial concentration, chemical treatment, etc.,
- Substratum properties: chemical composition, roughness, physical configuration, etc.,
- Hydrodynamics in the aqueous phase and possibly fluid-structure interaction if the substratum is not rigid.

Hence, the attachment and adhesion of biomass to the substratum depends on numerous parameters. When modelling attachment, this implies that many model parameters are required to capture this phenomenon. Including all of these different parameters in a model will likely make the model too complex. However, one possible mathematical simplification of this process is to consider the attachment phenomenon as a random event at each location of the substratum, where all the underlying mechanisms are lumped together into one driving stochastic process per location on the substratum. In other words, we view the attachment phenomenon as a random behaviour of the cells in the aqueous phase for which the underlying mechanisms of attachment are not considered. This notion is supported by experimental observations in [7, 8]. Thus, whatever processes may affect attachment at a given time and with a given set of system parameters are driven by a stochastic process at each inoculation point of the substratum.

The addition of biomass attachment has not readily been used in the modelling literature despite the importance of the process on biofilm growth and formation. Even when attachment is included in models, the approach is typically the addition of a simple deterministic function based on no

experimental data, which we discuss at the end of this section. The primary focus of this work is the mathematical modelling of discrete stochastic cell attachment and how this attachment may affect the spatial and temporal evolution of biofilms. The overall goal is to adapt the mathematical framework presented in [9]. The model presented in [9] has two major shortcomings, difficulty in model analysis and slow simulation times. The difficulty in model analysis comes from the Itô stochastic differential equations (SDEs) formalism [10], while the slow simulation times come from small step-sizes caused by the explicit method used. Therefore, we use random ordinary differential equations (RODEs), which allows us to use ODE theory and numerics path-wise. Moreover, RODEs allow to use different stochastic processes and these stochastic processes can appear in nonlinear terms (see Section 3.1).

Barring computational costs, we want our modelling framework to apply to biofilms of different types and morphologies and in any number of spatial dimensions. However, for analysis and simulation purposes, we apply our modelling framework to the case of 2D cellulolytic biofilms. Cellulolytic biofilms mainly arise in the biofuel production industry, and they have specific characteristics that we will exploit. The attachment phenomenon is assumed to be a discrete process and is characterized in terms of the adhesion factors discussed above; however, the attachment phenomenon is not very well-understood.

Biofuels are an alternative fuel source in the transportation sector in place of fossil fuels and can be implemented in the existing infrastructure. A significant benefit of biofuels is that it is an abundant renewable resource and has a net-zero carbon dioxide release into the atmosphere. However, the production of biofuels using food products, such as sugars, starch, fats, and oils, is not sustainable due to high and increasing demand for food [11]. Therefore, it is better to use inedible resources, such as lignocellulose*, to produce biofuels. The annual production of cellulosic ethanol in the US from lignocellulosic resources was estimated at 53 billion gallons in 2016 [12].

The production of cellulosic ethanol is considered achievable for lignocellulosic biomass. However, a few difficulties arise, including biomass recalcitrance, which is associated with the resistance of cell walls to enzymatic deconstruction, and the heterogeneity of the substrate caused by the complex structure of lignocellulose [13]. Therefore, the lignocellulose biomass needs to be treated and prepared before it can be extracted. In consolidated bioprocessing (CBP), cellulolytic bacteria perform fermentation and hydrolysis of the sugars in solid lignocellulosic substances [13, 14]. CBP requires micro-organisms that are highly efficient at fermenting the resultant sugars and producing cellulolytic enzymes. The micro-organisms should also produce little by-product and be able to live in environments with high levels of ethanol.

Common bacteria used in CBP include *Clostridium thermocellum* and *Caldicellulosiruptor obsidiansis*, which consume cellulosic substrate and form cellulolytic biofilms [7, 15]. These cellulolytic biofilm colonies are distributed spatially and evolve temporally [8]. Cellulolytic biofilms form very little EPS, attach directly to the substrate, and are typically monolayers [7, 8]. These cellulolytic biofilms differ from the traditional biofilms, which grow into the aqueous phase that contains the substrates. These cellulolytic biofilms carve out crater-like structures in the substratum, which are referred to as inverse colonies [8]. Cellulolytic biofilms form and grow through several stages (see Figure 1):

- 1) Initial cell attachment to the substratum,
- 2) Cell growth and division,

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

- 3) Inverted colony formation,
- 4) Emergence of crater-like structures due to substrate consumption and substratum degradation,
- 5) Radial growth of the crater-like depressions and the eventual coalescence of the crater-like structures,
- 6) Formation of a uniform-thickness biofilm.

The different stages of the spatiotemporal evolution of biofilms formed by *C. obsidiansis* on cellulose chads are captured by a confocal laser scanning microscope and are visualized in Figure 1 [8]. The crater-like structures that form are provided in Figure 1 [8].

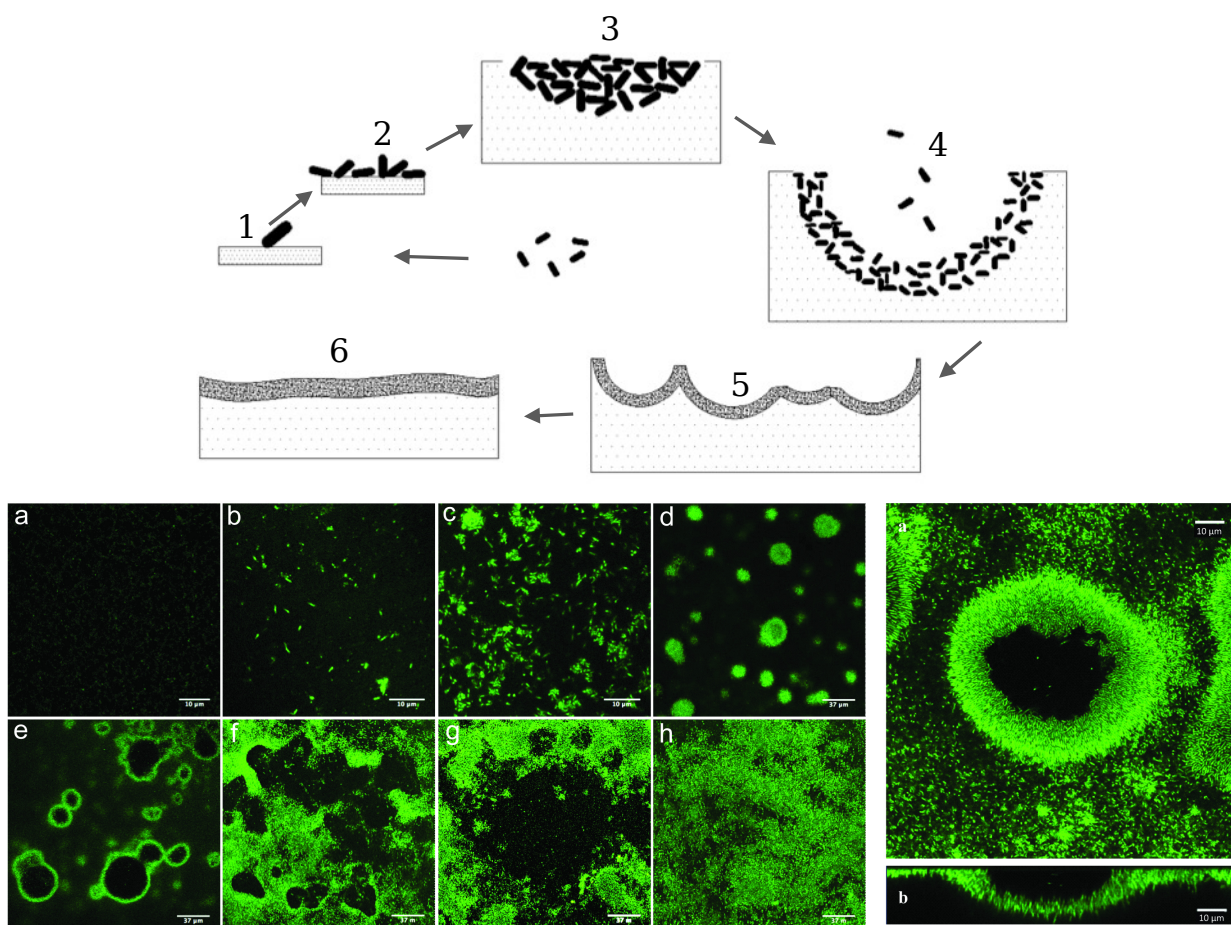


Figure 1. Top: Schematic diagram of the six stages of cellulolytic biofilm formation: 1) initial cell attachment, 2) cell growth and division, 3) inverted colony formation, 4) emergence of the crater-like formation, 5) further growth and merging of craters, 6) uniform biofilm formation, from [8]. **Bottom:** Confocal Laser Scanning Micrographs of cellulolytic biofilms. **Left:** 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. The radius of the craters is $\approx 40 \mu\text{m}$ [8]. **Right:** Crater formation by *C. thermocellum*. a) top view, b) cross-sectional view. Length scale = $10 \mu\text{m}$ [8]. All images are from [8].

The mathematical modelling of stochastic attachment of cells to cellulolytic biofilms has been

addressed in two different works. First, in an *ad hoc* manner by modifying a deterministic model of cellulolytic biofilms [16]. Secondly, [9] developed a model by first discretizing the deterministic model from [16] and then adding an additional stochastic term to the biomass equations using an attachment factor, which converted the model to an SDE system. Therefore, both models have the same deterministic part that deals with the formation and temporal evolution of cellulolytic biofilms and is described by a highly nonlinear degenerate coupled system of a partial differential equation (PDE) and an ODE for the biomass density and substrate concentration respectively. However, the SDE model in [9] is constructed via a spatial discretization of the PDE-ODE coupled system.

The model in [16] has an *ad hoc* inclusion of stochastic cell attachment via the addition of an impulse function to the biomass equation. The modelling mechanism behind the impulse function works as follows. If a certain quantity (Q) is greater than a uniform random number over $[0, 1]$, then 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 hence no biomass is added. The quantity Q is defined based on a few assumptions on the likelihood of an attachment event occurring. An attachment event is less likely to occur if the location has a large quantity of biomass, M , already, or if the substrate concentration, C , is too low, i.e., $Q \propto (1 - M) \times C$, where we assume that M is bounded by 1. By formulating the attachment phenomenon in this way, paper [16] was able to reproduce the qualitative behaviour of cellulolytic biofilms with cell attachment. This can be seen from the graphical visualization taken from [16] (see Figure 2) and the comparison with the experimental observations given in Figure 1.

Although the approach taken in [16] reproduces the expected qualitative behaviour, there is no mathematical framework and it is not based on a stochastic process. In particular, the model cannot be written as a stochastic or random (partial) differential equation.

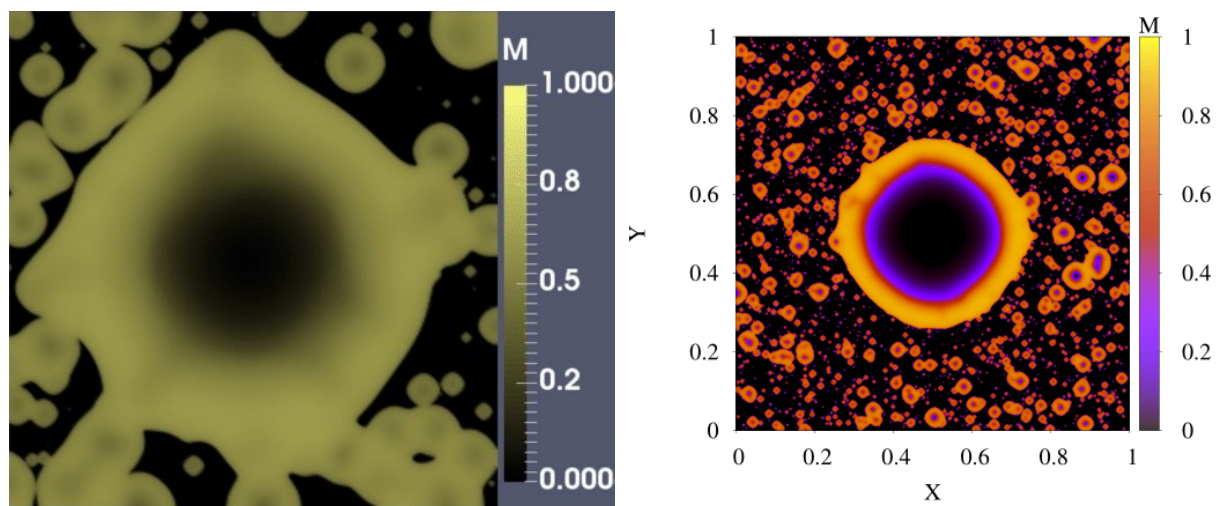


Figure 2. Example simulations of cellulolytic biofilm attachment models with a starting central inoculation, where $(0 \leq M \leq 1)$ stands for the biomass. **Left:** The PDE-ODE coupled system with random attachment presented in [16]. **Right:** An example simulation of the SDE random attachment model presented in [9].

The model in [9] seeks to utilize the insight gained from [16] but using a rigorous mathematical framework. In order to capture cellular attachment a similar impulse term is mollified and added to the biomass equation and an attachment factor is introduced for each grid cell of the domain. The attachment factor is a stochastic process based on the construction of Q given above, without drift. The attachment factor was simulated alongside the biomass and substrate equations. The model consists of an Itô SDE with the necessary regularity properties to apply SDE theory. The model in [9] reproduced the qualitative behaviour observed in experiments (see Figure 2). However, there are still significant downsides to this approach. The first being the similar *ad hoc* approach with an attachment factor, and the second being the numerical implementation. The numerical method for solving this model was the simplest available. Namely, [9] used the Euler-Maruyama method with a time-stepping approach constructed to ensure stability of the system. The downside of this method stems from the fact that discretized diffusion-reaction equations are typically stiff [17], especially with the nonlinear diffusion coefficient in this model. Therefore, the time-step required to maintain stability is quite small.

Other attachment modelling strategies are used in the literature for different types of biofilms. The approaches taken in other models are deterministic. For example, in [18], attachment was added to a 3D biofilm model via a pre-determined function of time. In other models, the attachment of biomass is prescribed as a function of the biomass concentration in the aqueous phase [19–21]. Typically the function αu , where α is the attachment rate and u is the concentration of biomass in the aqueous phase, is used. The attachment terms in both cases are not related to any underlying mechanism. This is similar to the approaches for cellulolytic biofilms [9, 16] and to the approach we will present. However, we take into account the stochastic nature of attachment, which might be a more realistic approach.

2. The spatially explicit deterministic continuum model

2.1. Model assumptions

The model in [16] adapts the model for traditional biofilms proposed in [22] to the case of cellulolytic biofilms. The aqueous phase does not explicitly appear in the model formulation and the nutrients are contained in the substratum. The model assumptions are given below.

- 1) The local capacity of the substratum for accommodating biomass is limited.
- 2) The only growth-limiting substrate is carbon. The carbon is only present inside the substratum and is assumed to be immobile and not replenished after consumption.
- 3) The spatial distribution, production, and movement of biomass only depends on the local availability of space and nutrients in the substratum. The biofilm does not exhibit notable expansion if there is available space.
- 4) Carbon is consumed by the bacteria and converted into new cells. We assume that biomass growth and carbon degradation are governed by standard saturation kinetics.
- 5) Cells are lost from the biofilm either by detachment or cell lysis. Both processes are combined into one process and are proportional to the local biomass density. Once carbon becomes depleted, biofilm growth ceases, and the loss of cells becomes the dominant process.
- 6) No cellular attachment occurs from the aqueous phase.

We will describe the two-dimensional variation of the model presented in [16]. More specifically, we will be considering the limiting case of cellulosic biofilms that spread across a very thin cellulosic

substratum, where the depth of the substratum is negligible compared to the length scales of the colonies. This is one of the two cases described in [16].

2.2. Governing equations

The model is a coupled system of a PDE describing the spatiotemporal evolution of the biomass density $M = M(t, x)$ and an ODE describing the evolution of the carbon concentration $C = C(t, x)$. We formulate the model in a rectangular spatial domain $\Omega := [0, 1] \times [0, W] \subset \mathbb{R}^2$, with $W \leq 1$. All variables and parameters are dimensionless (see [9] for more details). The model is given by

$$\partial_t M = \nabla \cdot (D(M)\nabla M) + F(C)M, \quad (t, x) \in (0, \infty) \times \Omega, \quad (2.1)$$

$$\partial_t C = -G(C)M, \quad (t, x) \in (0, \infty) \times \Omega. \quad (2.2)$$

Here $D(M)$ is the biomass diffusion coefficient, $F(C)$ is the net biomass growth rate, and $G(C)$ is the uptake rate of the carbon substrate. The substrate uptake rate is the only driving factor in (2.2) since there is no transport of carbon by model assumption 2.1. This is in contrast to the modelling framework presented in [22]. The actual biofilm consists of the sub-region of the domain Ω defined by

$$\Omega_M(t) := \{x \in \Omega \mid M(t, x) > 0\}.$$

Following [9, 16, 22], we define the biomass density dependent diffusion coefficient by

$$D(M) := \delta \frac{M^\alpha}{(1 - M)^\beta}, \quad \alpha, \beta \geq 1, \quad (2.3)$$

where $\delta > 0$ is the motility coefficient. By model assumption 2.1 and the non-dimensionalization given in [9], we assume that the maximum attainable biomass is 1.

The diffusion coefficient exhibits two distinct cases of behaviour: (i) if the local biomass density is low, then little diffusion will occur, meaning the substratum can accommodate additional biomass (in accordance with model assumption 2.1); (ii) if the local biomass density is close to its maximum value, then the equation for biomass will exhibit super diffusion behaviour and the biomass will move quickly to neighbouring locations (in accordance with model assumption 2.1). For $0 \leq M \ll 1$, the diffusion coefficient can be approximated by a degenerate power law, which is known from the porous medium equation, i.e., $D(M) \approx M^\alpha$, which approaches 0 as $M \rightarrow 0$. From simulations in [23], we know that if the initial data has compact support, then so does the solution. Therefore, the degenerate power-law creates an expanding biofilm colony with a sharp front. The exponent α in the diffusion coefficient controls how much biomass can be accommodated locally before a noticeable expansion into neighbouring regions occurs. When $0 \ll M < 1$, the biomass density approaches the singularity in the diffusion coefficient and super diffusion dominates, i.e., $D(M) \approx (1 - M)^{-\beta}$. The exponent β determines the severity of the super diffusion. Both diffusion modes are required to ensure expansion with finite speed (due to the degenerate power-law) and to have the local volume filling effect (due to the super diffusion) [16].

The net biomass growth rate is composed of the biomass growth rate and the cell loss rate. By model assumption 2.1, the cell loss rate includes both cell lysis and detachment of viable bacteria into the aqueous phase. By model assumption 2.1, carbon is the only growth-limiting substrate in the

system, and by model assumption 2.1 the biomass growth rate is described by Monod kinetics [24]. Therefore, the net growth rate is given by

$$F(C) = \frac{C}{\kappa + C} - \lambda, \quad (2.4)$$

where $\kappa > 0$ is the half-saturation constant and $\lambda > 0$ is the cell loss rate. The maximum growth rate is equal to 1 by our non-dimensionalization.

The uptake rate of the carbon substrate is modelled using the standard Monod saturation kinetics [24] by model assumption 2.1 and is given by

$$G(C) = \Upsilon \frac{C}{\kappa + C}, \quad (2.5)$$

where $\Upsilon > 0$ is the maximum specific consumption rate.

Biomass is restricted to the domain Ω and it is thus natural to impose homogeneous Neumann conditions on (2.1), i.e.,

$$\partial_n M(t, x) = 0, \quad (t, x) \in [0, \infty) \times \partial\Omega, \quad (2.6)$$

where ∂_n denotes the outward normal derivative and $\partial\Omega$ the boundary of Ω . We also impose the initial conditions

$$M(0, x) = M_0(x), \quad C(0, x) = C_0(x), \quad x \in \Omega. \quad (2.7)$$

We assume that $M_0(x) > 0$ in small subsets of Ω and $M_0(x) = 0$ everywhere else and we assume $C_0(x) > 0$ in Ω . In general, $C_0(x)$ can depend on the spatial position x ; however, in most of our simulations we restrict ourselves to the case where $C_0(x) = C_\infty = \text{const}$ as in the experiments conducted in [7, 8, 25], which used homogeneous paper chads.

2.3. Spatial discretization

For simulation purposes and to introduce our attachment term, we spatially discretize the PDE-ODE coupled system. We place an $N \times K$ uniform grid \mathcal{N} onto the rectangular domain $\Omega := [0, 1] \times [0, W]$ with $W \leq 1$. We use the finite volume scheme as introduced in [26] and described in detail in [9]. This converts our PDE-ODE coupled system into $2NK$ ODEs for each grid cell (i, j) as

$$\frac{dM_{i,j}}{dt} = \frac{1}{\Delta x} \sum_{\sigma \in \mathcal{N}_{i,j}} J_\sigma + \left(\frac{C_{i,j}}{\kappa + C_{i,j}} - \lambda \right) M_{i,j}, \quad (2.8)$$

$$\frac{dC_{i,j}}{dt} = -\Upsilon \frac{C_{i,j}}{\kappa + C_{i,j}} M_{i,j}. \quad (2.9)$$

where $\mathcal{N}_{i,j} := \{(i+1, j), (i-1, j), (i, j+1), (i, j-1)\}$ is the set of neighbouring grid cells and

$$J_\sigma = \begin{cases} \frac{1}{2\Delta x} (D(M_\sigma) + D(M_{i,j})) (M_\sigma - M_{i,j}) & \text{for } \sigma \in \mathcal{N} \\ 0 & \text{for } \sigma \notin \mathcal{N}, \end{cases} \quad (2.10)$$

is the flux out of the grid cell (i, j) into grid cell σ , where $\Delta x = 1/N = W/K$ is the side length of the grid cells. We write the $2NK$ ODEs in matrix-vector notation as

$$\begin{aligned}\frac{d\mathbf{M}}{dt} &= \mathcal{D}(\mathbf{M})\mathbf{M} + \mathcal{F}(\mathbf{C})\mathbf{M}, \\ \frac{d\mathbf{C}}{dt} &= -\mathcal{G}(\mathbf{C})\mathbf{M},\end{aligned}\tag{2.11}$$

after ordering the grid cells by $\pi(i, j) = (i-1)K + j$. The matrices $\mathcal{D}(\mathbf{M})$, $\mathcal{F}(\mathbf{C})$ and $\mathcal{G}(\mathbf{C})$ are $NK \times NK$ and represent the diffusion, net biomass growth, and carbon uptake, respectively.

Remark 2.1. *The matrix $\mathcal{D}(\mathbf{M})$ is symmetric and weakly diagonally dominant with non-negative off-diagonal entries and non-positive diagonal entries [9]. Therefore, $\mathcal{D}(\mathbf{M})$ is at least negative semi-definite. The matrices $\mathcal{F}(\mathbf{C})$ and $\mathcal{G}(\mathbf{C})$ are diagonal matrices with diagonal entries $f_{pp} = f(C_p)$ and $g_{pp} = G(C_p)$. If $C_p = 0$, then $f_{pp} = -\lambda$, the natural cell loss rate, and $g_{pp} = 0$.*

In the next section, we introduce our new approach for modelling cellular attachment via RODEs.

3. A spatially explicit model with discrete random attachment events

3.1. Random ordinary differential equations and the Wiener process

Here we introduce random ordinary differential equations (RODEs), which we will use to model random attachment. We follow the description in [27].

Let $(S, \mathcal{F}, \mathbb{P})$ be a probability space, where S is the set of outcomes, \mathcal{F} is a σ -algebra on S , and \mathbb{P} is a probability measure, and let $\eta : [0, T] \times S \rightarrow \mathbb{R}^m$ be an \mathbb{R}^m -valued stochastic process with continuous sample paths. In addition, let $g : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^d$. A RODE in \mathbb{R}^d is an equation of the form

$$\frac{dx}{dt} = g(x, \eta_t(s)), \quad x \in \mathbb{R}^d,\tag{3.1}$$

that can be viewed as a non-autonomous ordinary differential equation

$$\frac{dx}{dt} = G_s(t, x) := g(x, \eta_t(s))\tag{3.2}$$

for almost every realization $s \in S$ [27]. Since RODEs are just non-autonomous ODEs, we can apply results from ODE theory path-wise (for a given realization $s \in S$) to gain insight into the dynamics of an RODE, as long as our G_s still satisfies the regularity requirements. The results in our case only require continuity in t , which our stochastic processes satisfy.

In fact, we will use the Wiener process W_t . The Wiener process is a Gaussian process with the following properties [28]:

- 1) $W_0 = 0$ almost surely.
- 2) W_t has independent increments, namely $W_{t_1} - W_{\tau_1}$ and $W_{t_2} - W_{\tau_2}$ are independent random variables for all $0 \leq \tau_1 < t_1 \leq \tau_2 < t_2$.
- 3) The increments $W_t - W_\tau$ are normally distributed with $E(W_t - W_\tau) = 0$ and $Var(W_t - W_\tau) = t - \tau$ for all $0 \leq \tau < t$.
- 4) The sample paths are continuous almost surely.

3.2. Mathematical formulation of random cell attachment

Similarly to [9], we add discrete random attachment events. Attachment is the process by which cells in the aqueous phase attach to either the substratum or the pre-existing biomass. In many controlled experiments, no bacteria are present at the inflow, and thus, bacteria are only present in the aqueous phase if they have been previously detached from the biofilm. This is an example of the experiments done in [7, 29], on which the models presented in [9, 16] are based. Similar to [9], we make the simplifying assumption that attachment and detachment can be modelled independently. This is supported by [7] as they found that the availability of cells in the aqueous phase was not a limiting factor for attachment. Therefore, we assume that there is an unlimited supply of available cells for attachment. We also assume that cells generally do not attach to areas that have low carbon concentration or areas with high biomass density. This is consistent with the assumptions in [9, 16].

We add a mollified impulse function to each biomass Eq (2.8) with a mutually independent Wiener process W_t as argument for each grid cell (i, j) . As in [9], we use the Fermi-Dirac distribution. It allows us to easily and independently control the location of the impulse, the height of the impulse, and how well the mollification approximates a true discontinuous impulse function. The Fermi-Dirac distribution with parameters a, b, σ is given by

$$f(\phi) = \frac{1}{e^{\frac{\phi-b}{\sigma}} + 1} - \frac{1}{e^{\frac{\phi-a}{\sigma}} + 1}. \quad (3.3)$$

The parameters $a < b$ determines the location of the impulse, while the parameter σ controls how close the approximation is to the true impulse function. In particular, as $\sigma \rightarrow 0$ the Fermi-Dirac distribution converges to an impulse function supported on $[a, b]$ turning on at $\phi = a$ and off at $\phi = b$.

We now add to each biomass equation the term

$$\gamma g(M_{i,j})h(C_{i,j})f(W_t^{(i,j)}), \quad (3.4)$$

where $0 < \gamma \ll 1$ controls the intensity of bacterial attachment per unit time and $W_t^{(i,j)}$ is an independent standard Wiener process for each grid cell (i, j) . Here we see that the stochastic processes appear in the nonlinear function f . The functions g and h satisfy the attachment assumptions depending on M and C , *i.e.*, $g(1) = 0 = h(0)$, both $g(0), h(1) \leq 1$, g is decreasing and h is increasing, and these functions are sufficiently smooth. As in [9], we take functions of the form

$$g(M) = \begin{cases} (1 - M)^{\nu_1}, & M < 1, \\ 0, & M \geq 1, \end{cases}, \quad h(C) = C^{\nu_2}. \quad (3.5)$$

The exponents $\nu_1, \nu_2 > 0$ are included for generality. Hence, from (2.8) we obtain

$$\frac{dM_{i,j}}{dt} = \frac{1}{\Delta x} \sum_{\sigma \in \mathcal{N}_{i,j}} J_\sigma + \left(\frac{C_{i,j}}{\kappa + C_{i,j}} - \lambda \right) M_{i,j} + \gamma g(M_{i,j})h(C_{i,j})f(W_t^{(i,j)}). \quad (3.6)$$

Using the same ordering as before, our model becomes a system of $2NK$ RODEs given by

$$\begin{aligned} \frac{d\mathbf{M}}{dt} &= \mathcal{D}(\mathbf{M})\mathbf{M} + \mathcal{F}(\mathbf{C})\mathbf{M} + \mathcal{T}(\mathbf{M}, \mathbf{C})\mathbf{F}(W_t), \\ \frac{d\mathbf{C}}{dt} &= -\mathcal{G}(\mathbf{C})\mathbf{M}, \end{aligned} \quad (3.7)$$

where $\mathcal{D}(\mathbf{M})$, $\mathcal{F}(\mathbf{C})$, and $\mathcal{G}(\mathbf{C})$ are the same as in (2.11). Moreover, $\mathcal{T}(\mathbf{M}, \mathbf{C})$ is an $NK \times NK$ matrix with entries calculated based on (3.4) and $\mathbf{F}(\mathbf{W}_t)$ is an NK vector with components equal to (3.3).

Remark 3.1. The matrix $\mathcal{T}(\mathbf{M}, \mathbf{C})$ is a diagonal matrix with diagonal entries $i_{pp} = \gamma g(M_p)h(C_p)$. These entries vanish if either $M_p = 1$ or $C_p = 0$ and attain their maximum value if both $M_p = 0$ and $C_p = 1$. Therefore, $0 \leq i_{pp} \ll 1$. The p th entry of \mathbf{F} is given by $F_p(\mathbf{W}_t) = f(W_t^p) = \left(e^{\frac{W_t^p - b}{\sigma}} + 1 \right)^{-1} - \left(e^{\frac{W_t^p - a}{\sigma}} + 1 \right)^{-1}$. Finally, by construction, $0 \leq F_p \leq 1$.

3.3. Model analysis of a regularized system

As in previous works [9, 26, 30], we regularize the diffusion coefficient and consider

$$D^\epsilon(M) = \begin{cases} 0, & M < 0, \\ \delta \frac{M^\alpha}{(1-M)^\beta}, & 0 \leq M \leq 1 - \epsilon, \\ \delta \epsilon^{-\beta} (1 - \epsilon)^\alpha, & M > 1 - \epsilon, \end{cases} \quad (3.8)$$

for $\epsilon > 0$ sufficiently small. This leads to the regularized, approximate system

$$\begin{aligned} \frac{d\mathbf{M}^\epsilon}{dt} &= \mathcal{D}^\epsilon(\mathbf{M}^\epsilon)\mathbf{M}^\epsilon + \mathcal{F}(\mathbf{C}^\epsilon)\mathbf{M}^\epsilon + \mathcal{T}(\mathbf{M}^\epsilon, \mathbf{C}^\epsilon)\mathbf{F}(\mathbf{W}_t), \\ \frac{d\mathbf{C}^\epsilon}{dt} &= -\mathcal{G}(\mathbf{C}^\epsilon)\mathbf{M}^\epsilon. \end{aligned} \quad (3.9)$$

We perform analysis pathwise, i.e., for a given realization of the Wiener processes. We will show solutions are non-negative and bounded using comparison theorems and the tangent criterion from [31]. We will also study the long-term behaviour of our system. We first state a lemma that follows from standard Calculus results.

Lemma 3.2. Consider the scalar ODE

$$\frac{du}{dt} = -au + \rho(t)$$

with $a > 0$, and $\rho(t) \geq 0$ and continuous for all $t \geq 0$ with $\rho(t) \rightarrow \rho_\infty$ as $t \rightarrow \infty$. Then,

$$\lim_{t \rightarrow \infty} u(t) = \frac{\rho_\infty}{a}.$$

Proposition 3.3. Let \mathbf{M}^ϵ and \mathbf{C}^ϵ be path-wise solutions to (3.9) and suppose that the initial data satisfy $\mathbf{0} \leq \mathbf{M}^\epsilon(0) \leq \zeta \mathbf{1}$ with $\zeta \in (0, 1)$ and $\mathbf{0} \leq \mathbf{C}^\epsilon(0) \leq \mathbf{1}$. Then for almost every realization of the Wiener processes there exists a unique solution and the corresponding solution satisfies

$$\begin{aligned} \mathbf{0} &\leq \mathbf{M}^\epsilon \leq \mu \mathbf{1} \quad \text{for all } t \geq 0, \\ \mathbf{0} &\leq \mathbf{C}^\epsilon \leq \mathbf{1} \quad \text{for all } t \geq 0, \end{aligned}$$

where the constant $\mu > 0$ depends on the initial value and model parameters.

Proof. To prove the non-negativity, we invoke the tangent criterion [31]. First we consider the substrate equation. From Remark 2.1, the matrix $\mathcal{G}(\mathbf{C}^\epsilon)$ is diagonal and thus the p th component of the substrate equation is

$$-(\mathcal{G}(\mathbf{C}^\epsilon)\mathbf{M}^\epsilon)_p = -\Upsilon \frac{C_p^\epsilon M_p^\epsilon}{\kappa + C_p^\epsilon}.$$

For $C_p^\epsilon = 0$, the above vanishes and since $-\mathcal{G}(\mathbf{C}^\epsilon)\mathbf{M}^\epsilon$ satisfies a Lipschitz condition, it follows from the tangent criterion that $\mathbf{C}^\epsilon \geq 0$ [31].

The equation for the biomass consists of three components. From Remarks 2.1 and 3.1 we have

- p th component of $\mathcal{F}(\mathbf{M}^\epsilon, \mathbf{C}^\epsilon)\mathbf{F}(\mathbf{W}_t)$

$$\text{if}_p = \gamma g(M_p^\epsilon) h(C_p^\epsilon) f(W_t^p),$$

- p th component of $\mathcal{F}(\mathbf{C}^\epsilon)\mathbf{M}^\epsilon$

$$\text{fm}_p = \left(\frac{C_p^\epsilon}{\kappa + C_p^\epsilon} - \lambda \right) M_p^\epsilon,$$

- p th component of $\mathcal{D}^\epsilon(\mathbf{M}^\epsilon)\mathbf{M}^\epsilon$

$$\begin{aligned} \text{dm}_p = & \frac{1}{2\Delta x^2} [(D^\epsilon(M_{p-K}^\epsilon) + D^\epsilon(M_p^\epsilon))M_{p-K}^\epsilon + (D^\epsilon(M_{p-1}^\epsilon) + D^\epsilon(M_p^\epsilon))M_{p-1}^\epsilon \\ & - (D^\epsilon(M_{p-K}^\epsilon) + D^\epsilon(M_{p-1}^\epsilon) + 4D^\epsilon(M_p^\epsilon) + D^\epsilon(M_{p+1}^\epsilon) + D^\epsilon(M_{p+K}^\epsilon))M_p^\epsilon \\ & + (D^\epsilon(M_{p+1}^\epsilon) + D^\epsilon(M_p^\epsilon))M_{p+1}^\epsilon + (D^\epsilon(M_{p+K}^\epsilon) + D^\epsilon(M_p^\epsilon))M_{p+K}^\epsilon], \end{aligned}$$

where $M_i^\epsilon = 0$, whenever the grid cell does not exist.

If $M_p^\epsilon = 0$, then $\text{if}_p \geq 0$, $\text{fm}_p = 0$, and

$$\text{dm}_p = \frac{1}{2\Delta x^2} [D^\epsilon(M_{p-M}^\epsilon)M_{p-M}^\epsilon + D^\epsilon(M_{p-1}^\epsilon)M_{p-1}^\epsilon + D^\epsilon(M_{p+1}^\epsilon)M_{p+1}^\epsilon + D^\epsilon(M_{p+M}^\epsilon)M_{p+M}^\epsilon] \geq 0$$

if $M_{p-M}^\epsilon, M_{p-1}^\epsilon, M_{p+1}^\epsilon, M_{p+M}^\epsilon$ are all non-negative. Therefore, \mathbf{M}^ϵ is non-negative.

The non-negativity of \mathbf{M}^ϵ and \mathbf{C}^ϵ now implies that \mathbf{C}^ϵ is bounded by 1, as \mathbf{C}^ϵ is non-increasing. We finish by proving the upper bound for \mathbf{M}^ϵ . Since $d\mathbf{C}^\epsilon/dt \leq 0$, $\mathbf{0} \leq \mathbf{C}^\epsilon \leq \mathbf{1}$, and $-\mathcal{G}(\mathbf{C}^\epsilon)\mathbf{M}^\epsilon$ satisfies a Lipschitz condition, we can conclude that

$$\lim_{t \rightarrow \infty} \frac{d\mathbf{C}^\epsilon}{dt} = 0. \quad (3.10)$$

Therefore, each component of $d\mathbf{C}^\epsilon/dt \leq 0$ must converge to 0. Let

$$\boldsymbol{\eta}(t) := \mathcal{G}(\mathbf{C}^\epsilon)\mathbf{M}^\epsilon = -\frac{d\mathbf{C}^\epsilon}{dt}, \quad (3.11)$$

and define $\boldsymbol{\eta}_1(t) := \mathbf{1}^T \boldsymbol{\eta}(t)$. By (3.10), we know $\boldsymbol{\eta}_1(t) \rightarrow 0$ as $t \rightarrow \infty$. Now we can consider the ODE system

$$\begin{aligned} \frac{d\hat{\mathbf{M}}^\epsilon}{dt} &= \mathcal{D}^\epsilon(\hat{\mathbf{M}}^\epsilon)\hat{\mathbf{M}}^\epsilon + \mathcal{F}(\hat{\mathbf{C}}^\epsilon)\hat{\mathbf{M}}^\epsilon + \gamma \mathbf{1}, \\ \frac{d\hat{\mathbf{C}}^\epsilon}{dt} &= -\mathcal{G}(\hat{\mathbf{C}}^\epsilon)\hat{\mathbf{M}}^\epsilon. \end{aligned} \quad (3.12)$$

Let $\mathbf{M}_{tot} = \sum_{p=1}^{NK} \hat{M}_p^\epsilon$. Then, we can construct a differential inequality for \mathbf{M}_{tot} by summing up all components of the \hat{M}^ϵ DE in (3.12). This gives

$$\frac{d\mathbf{M}_{tot}}{dt} = \sum_{p=1}^{NK} F(\hat{C}_p^\epsilon) \hat{M}_p^\epsilon + NK\gamma = \sum_{p=1}^{NK} \left[\frac{\hat{C}_p^\epsilon}{\kappa + \hat{C}_p^\epsilon} - \lambda \right] \hat{M}_p^\epsilon + NK\gamma \leq -\lambda \mathbf{M}_{tot} + \frac{\eta_1(t)}{\Upsilon} + NK\gamma. \quad (3.13)$$

From this we define the barrier ODE

$$\frac{d\bar{\mathbf{M}}_{tot}}{dt} + \lambda \bar{\mathbf{M}}_{tot} = \frac{\eta_1(t)}{\Upsilon} + NK\gamma, \quad (3.14)$$

which we know is an upper bound on \mathbf{M}_{tot} by comparison theorems [31]. By Lemma 3.2, we know

$$\lim_{t \rightarrow \infty} \bar{\mathbf{M}}_{tot}(t) = \frac{NK\gamma}{\lambda}. \quad (3.15)$$

The ODE (3.14), is a standard linear ODE, and thus the solution does not tend to infinity in finite time and since $\mathbf{M}_{tot} \geq 0$, we have

$$\bar{\mathbf{M}}_{tot} \leq \mu \quad (3.16)$$

for some $\mu \geq 0$ that depends on model parameters and initial data. Since

$$\text{if}_p = \gamma g(M_p^\epsilon) h(C_p^\epsilon) f(W_t^p) \leq \gamma, \quad (3.17)$$

we observe that the solutions to (3.12) are upper bounds for the solutions to our RODE model (3.9), which implies the biomass density is bounded.

Since the Wiener process is pathwise continuous and the right-hand side of (3.9) satisfies a local Lipschitz condition, we have a unique local solution, and by the boundedness condition, we have a unique global solution [31]. \square

We now move on to showing the long-term behaviour of (3.9).

Proposition 3.4. *All solutions to (3.7), with non-negative initial data, satisfy*

$$\lim_{t \rightarrow \infty} \mathbf{M}^\epsilon(t) = \mathbf{0} \quad \text{and} \quad \lim_{t \rightarrow \infty} \mathbf{C}^\epsilon(t) = \tilde{\mathbf{C}}, \quad (3.18)$$

where $\mathbf{0} \leq \tilde{\mathbf{C}} \leq \mathbf{1}$.

Proof. For notational simplicity, we denote by $\mathbf{M} = \mathbf{M}(t)$ and $\mathbf{C} = \mathbf{C}(t)$ the solutions to (3.9). By (3.10) in the previous proof

$$\lim_{t \rightarrow \infty} -\mathcal{G}(\mathbf{C})\mathbf{M} = \mathbf{0}. \quad (3.19)$$

We also know that

$$(\mathcal{G}(\mathbf{C})\mathbf{M})_p = \Upsilon \frac{C_p}{\kappa + C_p} M_p, \quad (3.20)$$

and hence, $\mathcal{G}(\mathbf{C})\mathbf{M} = \mathbf{0}$ if and only if $C_p M_p = 0$ for all $p = 1, \dots, NK$. Therefore, since we know \mathbf{M} is bounded by a constant, we have

$$(\mathbf{M}^T \mathcal{T}(\mathbf{M}, \mathbf{C})\mathbf{F}(\mathbf{W}_t))_p = \gamma g(M_p) M_p C_p^{v_2} f(W_t^p) \rightarrow 0 \quad (3.21)$$

as $t \rightarrow \infty$ for all p . Therefore, $\mathbf{M}^T \mathcal{G}(\mathbf{C})\mathbf{M} / \Upsilon + \mathbf{M}^T \mathcal{T}(\mathbf{M}, \mathbf{C})\mathbf{F}(\mathbf{W}_t) \rightarrow 0$ as $t \rightarrow \infty$. Since we have the existence of a unique solution to (3.9) by Proposition 3.3, we know $\mathbf{M}^T \mathcal{G}(\mathbf{C})\mathbf{M} / \Upsilon + \mathbf{M}^T \mathcal{T}(\mathbf{M}, \mathbf{C})\mathbf{F}(\mathbf{W}_t)$ is continuous. To this end we define

$$\eta(t) := \frac{1}{\Upsilon} \mathbf{M}^T \mathcal{G}(\mathbf{C})\mathbf{M} + \mathbf{M}^T \mathcal{T}(\mathbf{M}, \mathbf{C})\mathbf{F}(\mathbf{W}_t), \quad (3.22)$$

where we know $\eta(t)$ is continuous for all $t > 0$ and $\eta(t)$ converges to $\eta_\infty := 0$ as $t \rightarrow \infty$. We will now use a discrete energy estimate type argument to show die-out of biomass. Let

$$E(t) := \|\mathbf{M}(t)\|_2^2. \quad (3.23)$$

Then if we multiply the biomass equation in (3.9) by $2\mathbf{M}^T$ on the left we get

$$\frac{dE}{dt} = 2\mathbf{M}^T \frac{d\mathbf{M}}{dt} = 2\mathbf{M}^T \mathcal{D}_\epsilon(\mathbf{M})\mathbf{M} + 2\mathbf{M}^T \mathcal{G}(\mathbf{C})\mathbf{M} + 2\mathbf{M}^T \mathcal{T}(\mathbf{M}, \mathbf{C})\mathbf{F}(\mathbf{W}_t). \quad (3.24)$$

By Remark 2.1, we know $\mathcal{D}(\mathbf{M})$ is negative semi-definite. The same reasoning applies to the regularized diffusion matrix and so $\mathcal{D}_\epsilon(\mathbf{M})$ is also negative semi-definite. Therefore, $\mathbf{M}^T \mathcal{D}_\epsilon(\mathbf{M})\mathbf{M} \leq 0$. So,

$$\frac{dE}{dt} \leq -2\lambda E + 2\eta(t). \quad (3.25)$$

since $\mathcal{G}(\mathbf{C}) = \Upsilon(\mathcal{F}(\mathbf{C}) + \lambda I)$, where I is the identity matrix. From this we define the barrier ODE

$$\frac{d\bar{E}}{dt} + 2\lambda \bar{E} = 2\eta(t), \quad (3.26)$$

which we know is an upper bound on (3.24) by comparison theorems [31]. By Lemma 3.2, we know that

$$\lim_{t \rightarrow \infty} \bar{E}(t) = \frac{\eta_\infty}{\lambda} = 0. \quad (3.27)$$

Since, by definition, E is non-negative, we know that $E \rightarrow 0$ as $t \rightarrow \infty$. Therefore, $\mathbf{M}(t) \rightarrow 0$ as $t \rightarrow \infty$ since $E(t) = \|\mathbf{M}(t)\|_2^2 = 0$ if and only if $\mathbf{M}(t) = 0$. \square

Remark 3.5. *We cannot prove that the biomass is strictly bounded by 1, which is an underlying assumption of the model. For generic parameters and initial data, it can occur that $M \geq 1$ in finite time. This issue is due to the homogeneous Neumann boundary conditions, which imply that no biomass can leave the domain and biomass decays if and only if C is sufficiently small. We can determine the critical substrate concentration for a given grid cell, such that biomass loss will start dominating locally. This occurs when $C < \lambda\kappa/(1 - \lambda) =: C_{crit}$. We always assume $\lambda < 1$; otherwise, cell loss would always dominate growth. Strictly speaking, our model is only valid as long as $M < 1$. However, during all simulations that we will present, the biomass concentration never reached 1, i.e. C dropped below C_{crit} before M could reach unity. Therefore, the restriction of our model to the case where $M < 1$ has no practical effect.*

4. Computational implementation

For numerical integration of the RODE system (3.7) we use the ODE trapezoid method with constant time-stepping. We use ODE numerics as the traditional SDE and RODE numerics [32, 33] led to instability, negative values or excessive simulation runtime, and were not remedied by the techniques found in [34]. To solve the non-linear system of equations, we use a fixed-point iteration scheme to convert the non-linear system into a sequence of linear systems. Namely, let U be either state variable, then U^k denotes the solution at the current time step and U^{k+1} the solution at the next time step. We denote by $U^{(p)}$ and $U^{(p+1)}$ the current and next iterate in our fixed point iteration in the current timestep. To simplify notation we omit here the time step counter k . Then, in each time step we iterate

$$\left[I - \frac{1}{2} \left(\Delta t \mathcal{D}(\mathbf{M}^{(p)}) + \Delta t \mathcal{F}(\mathbf{C}^{(p)}) \right) \right] \mathbf{M}^{(p+1)} = \mathbf{M}^k + \frac{\Delta t}{2} \left(\mathcal{T}(\mathbf{M}^{(p)}, \mathbf{C}^{(p)}) \mathbf{F}(\mathbf{W}_{t_{k+1}}) \right. \\ \left. + \mathcal{D}(\mathbf{M}^k) \mathbf{M}^k + \mathcal{F}(\mathbf{C}^k) \mathbf{M}^k \right. \\ \left. + \mathcal{T}(\mathbf{M}^k, \mathbf{C}^k) \mathbf{F}(\mathbf{W}_{t_k}) \right), \quad (4.1)$$

$$\mathbf{C}^{(p+1)} = \mathbf{C}^k - \frac{\Delta t}{2} \left(\mathcal{G}(\mathbf{C}^{(p+1)}) \mathbf{M}^{(p+1)} + \mathcal{G}(\mathbf{C}^k) \mathbf{M}^k \right), \quad (4.2)$$

over p until the 1–norm of successive iterations is less than some pre-specified tolerance. For our simulations we use a tolerance of 10^{-6} . Once convergence is reached we set $\mathbf{M}^{k+1} = \mathbf{M}^{(p+1)}$ and $\mathbf{C}^{k+1} = \mathbf{C}^{(p+1)}$. Each linear system for $\mathbf{M}^{(p+1)}$ is solved using the Conjugate Gradient method [35], since the resulting system is symmetric positive definite [36]. See [36] for more details.

Simulations were conducted using Julia 1.5.3 [37]. To simulate the independent Wiener processes, we use *randn*, which is a Gaussian random number generator built into Julia. Namely, given the previous value of the Wiener process W_{t_k} at time t_k , we calculate the next value $W_{t_{k+1}}$ at time t_{k+1} , via $W_{t_{k+1}} = W_{t_k} + \Delta W_{t_k}$, where ΔW_{t_k} is a Gaussian random number with mean 0 and variance $\Delta t = t_{k+1} - t_k$. We control the seeding of the random number generation by the *Random.seed!* function.

We parallelize the construction of the linear systems, the pseudo-random number generation (PRNG), the array copying, and the convergence calculations. We used two different parallelization techniques in Julia. The first is the macro *@threads*, which is placed in front of for loops, while the second is the *vmapntt* function, which provides vectorized parallel execution of a scalar function. For both methods, the ‘-t n’ runtime flag was used to control the number of threads available, where n is the number of threads. For more details see [36].

We ran all of our simulations on two separate computers. The first is a custom-built workstation with an AMD Ryzen Threadripper 3955WX (3.9 GHz, 16 cores, 32 threads) and 64 GB RAM running Ubuntu 20.10 (Groovy Gorilla). The second workstation is a custom-built Intel Xeon E5-1660 v4 (3.2 GHz, 6 cores, 12 threads) with 32 GB RAM running Ubuntu 16.04.7 LTS (Xenial Xerus).

5. Simulation results and computational analysis

5.1. 2D simulations

We begin by showcasing several simulations to illustrate the efficacy of our model and numerical method. The first set is with an initially clean domain, i.e., the initial biomass density is zero

everywhere. The second set investigates the model with an initially colonized domain, which starts with a small circular inoculation of biomass to see how the random attachment events interact with pre-existing biomass.

Table 1. Model parameters used for the simulations of (3.7). All values in the table are dimensionless.

Parameter	Symbol	Value	Reference
Motility Coefficient	$\tilde{\delta}$	10^{-6}	[16]
Diffusion Coefficient Exponent 1	α	4.0	[16]
Diffusion Coefficient Exponent 2	β	4.0	[16]
Maximum Consumption Rate	$\tilde{\Upsilon}$	0.4	[16]
Half Saturation Concentration	$\tilde{\kappa}$	0.01	[16]
Cell Loss Rate	$\tilde{\lambda}$	0.42	[16]
Control Parameter of f	σ	0.000001	assumed
Shift Parameters of f	a & b	Varied	assumed
Density of an Attached Cell per Time Step	γ	0.01	assumed
Biomass Attachment Exponents	$\nu_{1/2}$	2.0	assumed
Domain	Ω	$[0, 1] \times [0, 1]$	assumed
Grid Size	$N \times K$	$2^9 \times 2^9$	assumed
Step Size	Δt	10^{-2}	assumed
Seed	–	1	assumed

5.1.1. Case 1: initially clean domain

Here we simulate (3.7) with the initial conditions $M_0 \equiv 0$ and $C_0 \equiv 1$ everywhere in Ω . Between the two simulations, we controlled the emerging characteristics of the biomass by the impulse interval $[a, b]$. Both simulations use the parameter values given in Table 1 with Simulation 1 having an attachment interval $[a, b] = [0.99999, 1.0]$ (see Figure 3) while Simulation 2 has an attachment interval $[a, b] = [0.9999, 1.0]$ (see Figure 4). As the simulations progress temporally, we see small colonies form across the domain; in both simulations, this occurs at around $t = 20$. These colonies grow rapidly, and once they reach saturation, they expand over to neighbouring colonies. The simulations manifest the crater-like structures (see Figure 1) resembling the colonies formed in cellulolytic biofilms (see Figure 1). As the colonies continue to spread, some colonies coalesce and form communities. These communities continue to coalesce and form until the substratum is fully degraded. Once the substrate in the substratum becomes limited, the biomass begins to decay until no substrate is present and, in turn, no biomass.

Figures 3 and 4 show subtle differences in the colony formation. In Figure 3, we have fewer attachment events taking place, which is expected due to the smaller impulse interval $[a, b]$. At time $t = 20$ we have slightly more established attachment events in Figure 4 compared to Figure 3. If an attachment event occurs in the simulation depicted in Figure 3, then it must occur in the simulation depicted in Figure 4 since $[0.99999, 1] \subset [0.9999, 1]$ and Wiener processes have continuous sample paths. The simulation ends earlier for the larger impulse interval. Meaning the substrate and biomass reaching zero (or close to 0) everywhere. On Figure 3 we see that the simulation finishes shortly after

$t = 50$, while in Figure 4 it finishes at $t = 45$. This is caused by the slightly sooner attachment events and the larger total biomass resulting from more attachment events. More biomass in the system, especially spread out across more grid cells, means quicker substrate degradation. More substrate degradation means faster biomass death and detachment, as discussed in Remark 3.5.

Lastly, although it is difficult to see in these visualisations, the attachment events that contribute to substantial additional biomass, i.e. when $\gamma g(M)h(C)f(W_t) \approx \gamma$, cease in the grid locations that have high biomass density or low substrate concentration.

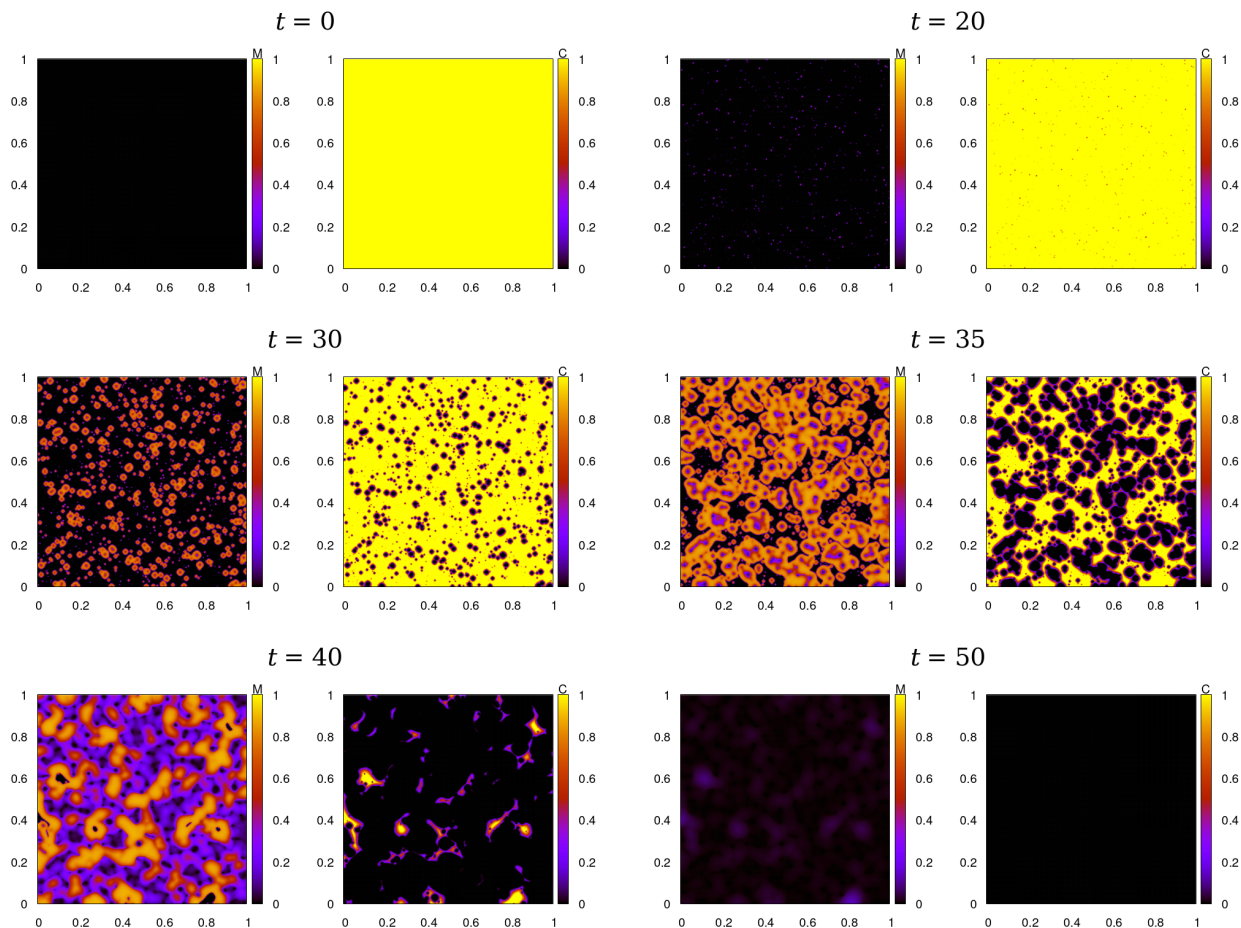


Figure 3. Simulation of (3.7) with initial data $M_0 \equiv 0$ and $C_0 \equiv 1$ everywhere. The biomass densities (left) with their corresponding substrate concentrations (right) are provided at the different time points listed. The interval of attachment is $[0.99999, 1.0]$.

5.1.2. Case 2: initially colonized domain

Here we simulated (3.7) with the initial biomass density given by

$$M_0(x, y) = \begin{cases} -\frac{h}{d^2} \left((x - 0.5)^2 + (y - 0.5)^2 \right) + h, & \text{if } (x - 0.5)^2 + (y - 0.5)^2 < d^2, \\ 0 & \text{otherwise,} \end{cases} \quad (5.1)$$

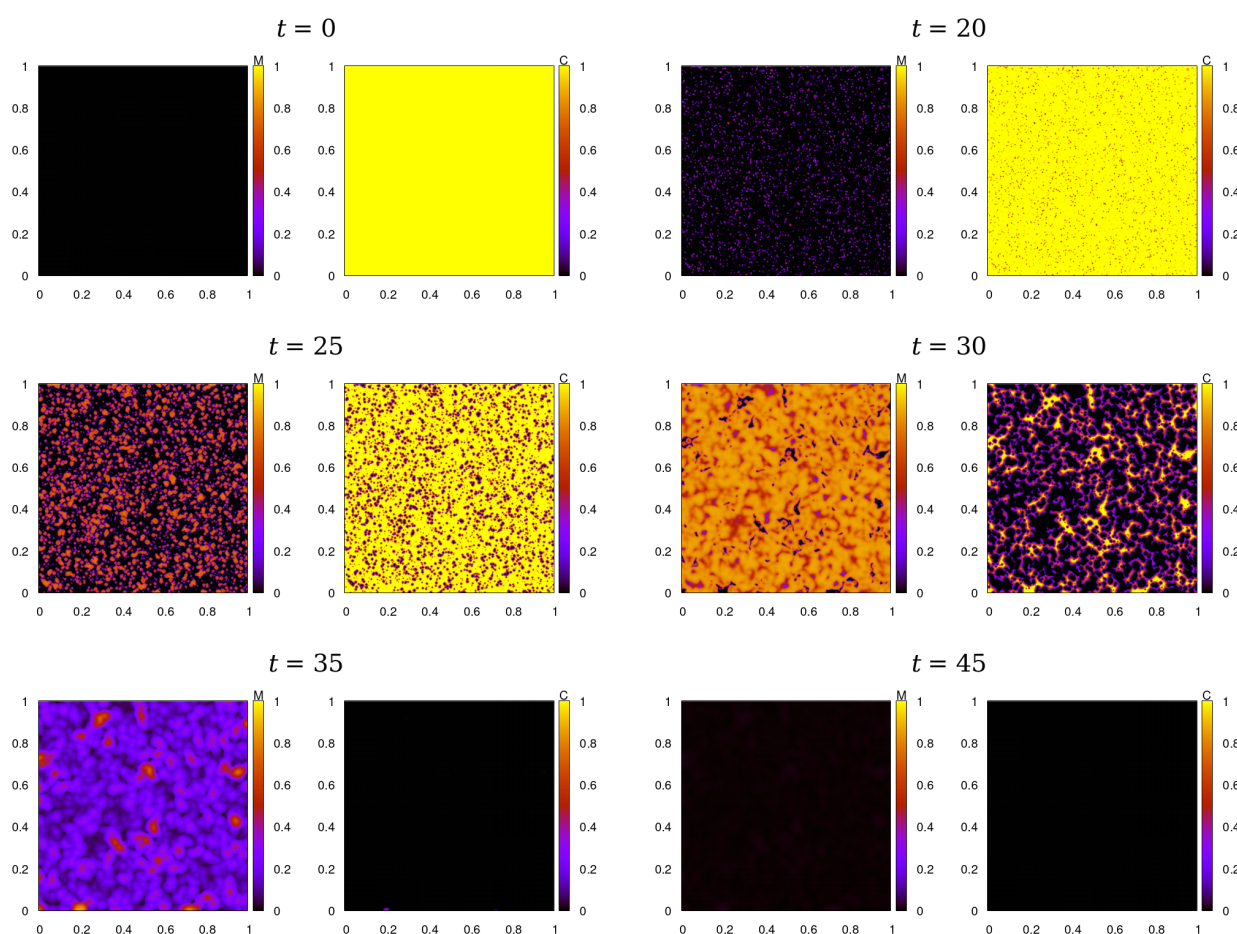


Figure 4. Simulation of (3.7) with initial data $M_0 \equiv 0$ and $C_0 \equiv 1$ everywhere. The biomass densities (left) with their corresponding substrate concentrations (right) are provided at the different time points listed. The interval of attachment is $[0.9999, 1.0]$.

with $h = 0.1$, $d = 5/127$, and $C_0 \equiv 1$ everywhere in Ω . The goal of these simulations is to examine the effects of attachment when a substantial bacterial colony is already present in the system.

We present two separate simulations. Both using the default simulation parameters in Table 1 with attachment intervals $[a, b] = [0.99999, 1.0]$ (see Figure 5) and $[a, b] = [0.9999, 1]$ (see Figure 6).

From Figures 5 and 6 we see that before attachment events start taking place, the central colony expands radially outward. The characteristic for cellulolytic biofilms craterlike depressions form. Initially they have a rim of approximately constant thickness in the wake of which eventually all biomass decays upon depletion of all nutrients. Eventually after this crater like structure merges with other colonies, the biomass vanishes as it attempts to expand into regions in which nutrients already have been depleted, i.e., into regions in which biomass loss terms dominate over growth. This is in agreement with experimental observations reported in [25] based on microscopy and online CO_2 measurements that monitor microbial activity. Already the deterministic model in [16], on which our work is based, and the SDE model in [9], upon which we aim to improve, showed this behavior. At

around $t = 20$ in both simulations, we see attached colonies start to form, which is consistent with the initially clean domain simulations as we use the same seeding for the stochastic processes. However, unlike the previous simulations, we see that no attachment occurs on or inside the central colony, which is caused by the lack of substrate and/or high biomass densities at these locations, and therefore unfavorable conditions for attachment per our model assumptions.

Like in the previous simulations, for larger impulse intervals, more attachment events occur, and when the impulse interval is closer to 0, attachment events occur slightly sooner. In both cases, the colonies that form due to attachment have ample time to grow and expand before they merge with the central colony. They consume the majority of the substrate in the sections of the domain which they occupy. As a result, the central colony cannot reach the edges of the domain before becoming completely distorted by the attached biomass. Unsurprisingly, the simulation with the larger attachment interval (see Figure 6) consumes the substrate quicker, distorts the central colony much sooner and to a greater degree when compared to the simulation with the smaller attachment interval (see Figure 5).

When comparing our simulations illustrated in Figures 3–6 to the empirical imaging in Figure 1 we see that Figures 3 and 5, which both use the attachment interval $[a, b] = [0.99999, 1.0]$, best capture the amount of biomass deposited onto these domains.

5.2. Grid refinement study

We performed a spatial grid refinement study to investigate the variations in our attachment model (3.7) as we refine the grid. Due to the stochastic nature of the model, we ran an ensemble of simulations for each grid size. We use the generally accepted notion that 30 samples are sufficient to apply the Central Limit Theorem[†] and thus, we perform 30 simulations for each grid size [38]. To compare the results between the grid resolutions, we keep the size of the attachment site constant. This preserves the added energy in the system.

We generate two statistics, the sample mean (expectation) and the standard error of the sample mean, given by

$$\text{sample mean} := E[X] = \frac{1}{n} \sum_{i=1}^n X_i, \quad (5.2)$$

$$\text{standard error of the sample mean} := SEM[X] = \frac{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - E[X])^2}}{\sqrt{n}}. \quad (5.3)$$

where $X = (X_1, \dots, X_n)$ denotes the samples and n is the number of realizations (samples).

The quantities of interest (QI) we will use are the total biomass and carbon,

$$M_{TOT}(t) := \int_{\Omega} M(t, x) dx = \Delta x^2 \sum_{p=1}^{NK} M_p(t), \quad (5.4)$$

$$C_{TOT}(t) := \int_{\Omega} C(t, x) dx = \Delta x^2 \sum_{p=1}^{NK} C_p(t), \quad (5.5)$$

[†]The use of the Central Limit Theorem will become more apparent when we discuss the stability of travelling wave solutions. We kept the number of simulations the same across different experiments for consistency sake.

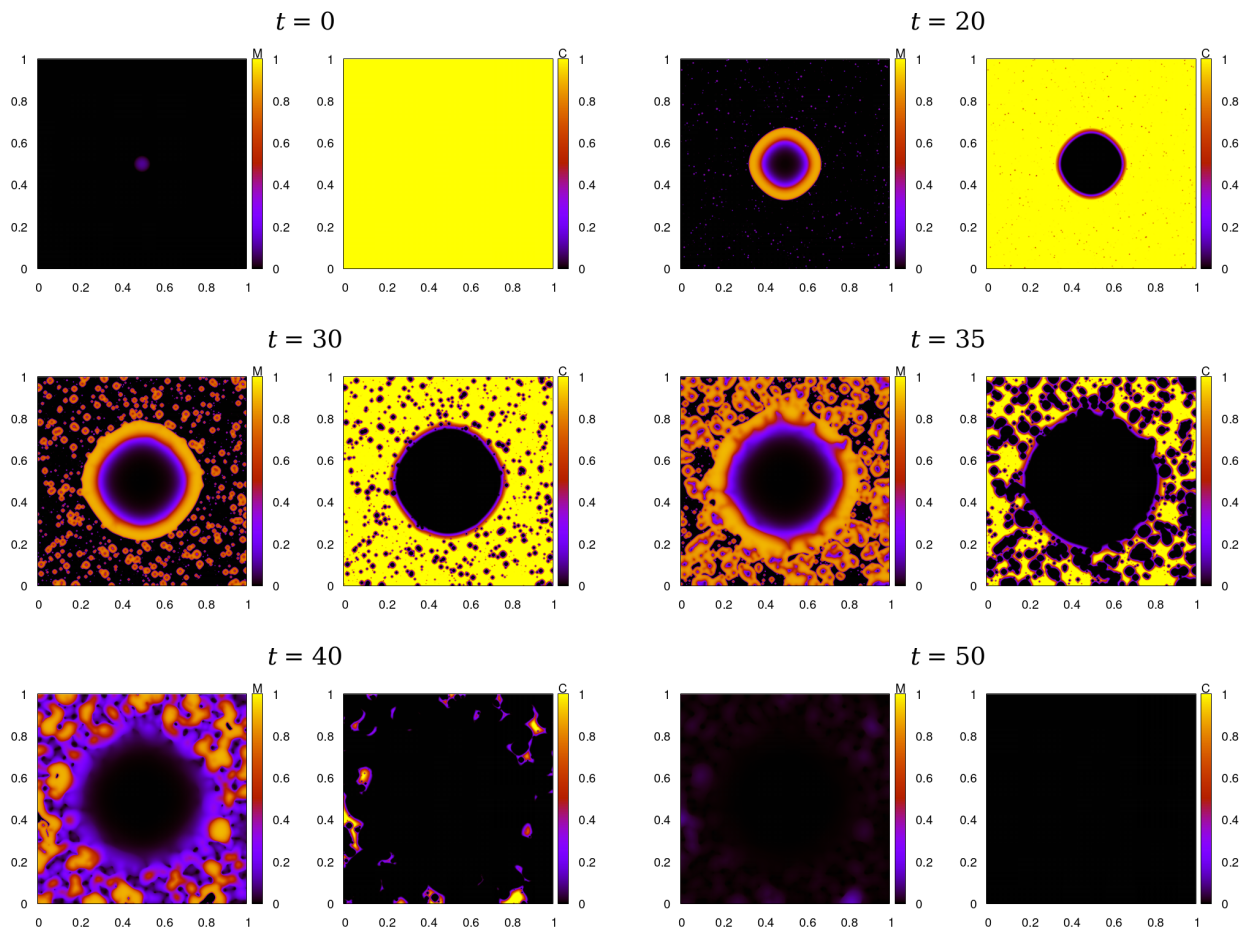


Figure 5. Simulation of (3.7) with initial data M_0 given by (5.1) and $C_0 \equiv 1$. The biomass densities (left) with their corresponding substrate concentrations (right) are provided at the different time points listed. The interval of attachment is $[0.99999, 1.0]$.

where Ω is the spatial domain, NK the number of grid cells, and $M_p(t)$, $C_p(t)$ denote the p th component of the vectors $\mathbf{M}(t)$ and $\mathbf{C}(t)$. The occupancy of the biomass and carbon,

$$M_A(t) := \int_{\Omega'_M(t)} dx = \Delta x^2 \sum_{p=1}^{NK} U_p^M(t), \quad (5.6)$$

$$C_A(t) := \int_{\Omega'_C(t)} dx = \Delta x^2 \sum_{p=1}^{NK} U_p^C(t), \quad (5.7)$$

where $\Omega'_M(t) := \{x \in \Omega \mid M(t, x) > 10^{-2}\}$ and $\Omega'_C(t) := \{x \in \Omega \mid C(t, x) > 10^{-2}\}$, and the points $U_p^L(t)$ for $L = M, C$ are 1 if $L_p(t) > 10^{-2}$ and zero otherwise. We use 10^{-2} as an approximation to 0 to avoid numerical noise. Lastly, we investigate the number of distinct biofilm colonies and depleted substrate colonies, which we denote by $M_{CC} = M_{CC}(t)$ and $C_{CC} = C_{CC}(t)$ for biomass and carbon respectively. To calculate the number of different colonies we determine the number of connected components using

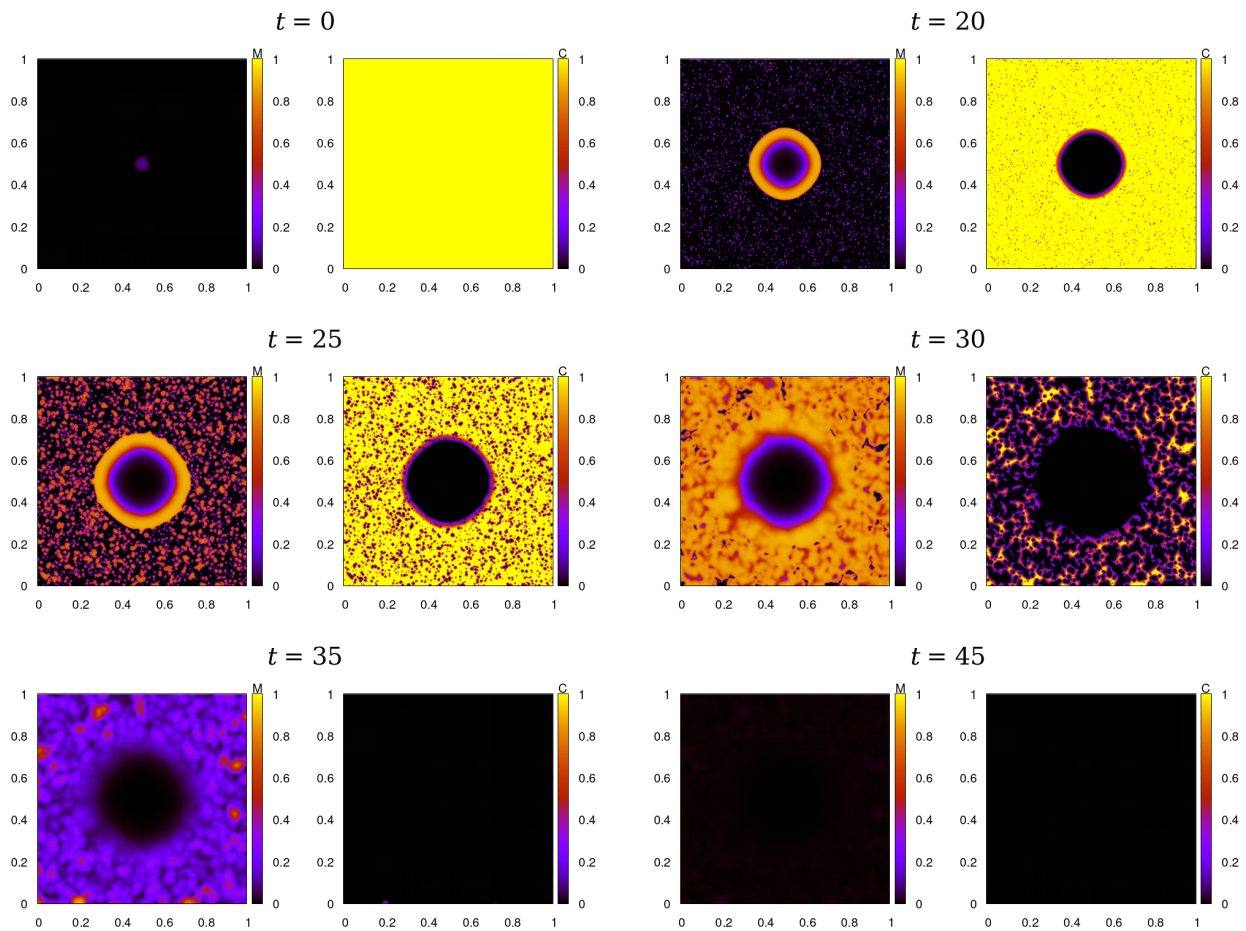


Figure 6. Simulation of (3.7) with initial data M_0 given by (5.1) and $C_0 \equiv 1$. The biomass densities (left) with their corresponding substrate concentrations (right) are provided at the different time points listed. The interval of attachment is $[0.9999, 1.0]$.

$\Omega'_{M/C}(t)$ with the near-nearest neighbour formulation [39].

We also calculate the relative distance between the expectation curves given by

$$dist_N = \frac{\|U_N - U_{1024}\|_p}{\|U_{1024}\|_p} \quad (5.8)$$

with $p = 1, 2, \infty$ denoting the usual p -norms and U_N denoting the expectation time series for the $N \times N$ grid.

For all grid refinement simulations, we used the default parameter values given in Table 1 and an attachment interval of $[a, b] = [0.99, 1.0]$. We use a larger attachment interval for the grid refinement study since the number of Wiener processes is less. We set the initial conditions to $M \equiv 0$ and $C \equiv 1$ in order to simulate an initially clean domain.

The time series of the sample means of the total biomass and substrate, the occupancy, and the number of communities are given in Figures 7, 9 and 11 with the associated standard error plots given in Figures 8, 10 and 12. The relative distances between the grid sizes are given in Table 2.

Table 2. Relative distance between the expectations of the three quantities of interest for various grid resolutions. All distances are calculated relative to the 1024×1024 grid.

Grid Resolution	Biomass Density (M)			Carbon Concentration (C)		
	$\ \cdot\ _1$	$\ \cdot\ _2$	$\ \cdot\ _\infty$	$\ \cdot\ _1$	$\ \cdot\ _2$	$\ \cdot\ _\infty$
Total Mass						
64×64	0.0757	0.0861	0.1135	0.01	0.0222	0.0648
128×128	0.0491	0.0542	0.068	0.0067	0.0144	0.0401
256×256	0.0255	0.0277	0.0338	0.0036	0.0075	0.0204
512×512	0.0095	0.0102	0.0123	0.0014	0.0028	0.0075
Occupancy						
64×64	0.0508	0.0932	0.2535	0.0125	0.0309	0.107
128×128	0.0328	0.058	0.1483	0.0081	0.0187	0.0606
256×256	0.0174	0.0299	0.0692	0.0042	0.0094	0.0294
512×512	0.0067	0.0113	0.0249	0.0016	0.0035	0.0106
Communities						
64×64	0.0853	0.1148	0.1761	0.577	0.6319	0.7099
128×128	0.0526	0.0763	0.1179	0.4011	0.4203	0.4707
256×256	0.0274	0.0417	0.0651	0.2106	0.2175	0.2442
512×512	0.0111	0.0174	0.0282	0.0829	0.0822	0.0916

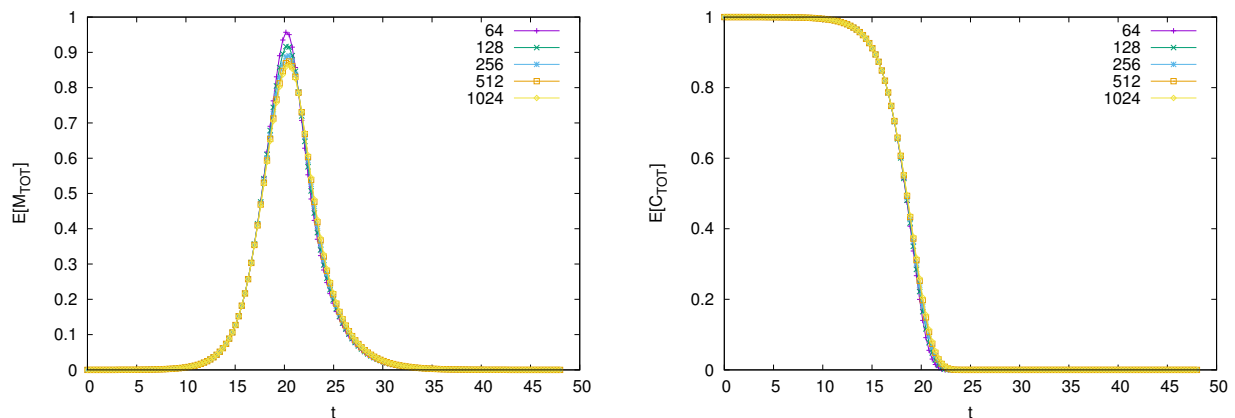


Figure 7. The time evolution of the expectation of total biomass (left) and substrate (right) for several grid sizes averaged over 30 realizations. The spatial size of an attachment event is $\frac{1}{64} \times \frac{1}{64}$ of the domain.

We see that for finer grid resolutions, the mean of the total biomass reaches a smaller peak value and the mean of the total biomass and carbon decay slower as a result, which is similar to the decay of biomass and carbon occupancy. This is interesting since Δx only appears in the diffusion coefficient, which plays no role for the total biomass. We also see that the mean biomass occupancy increases slower as the grid is refined. The smaller peaks and slower increase of the total mass and occupancy plots are in contrast to the colonies plots where we see higher biomass and depleted carbon colonies

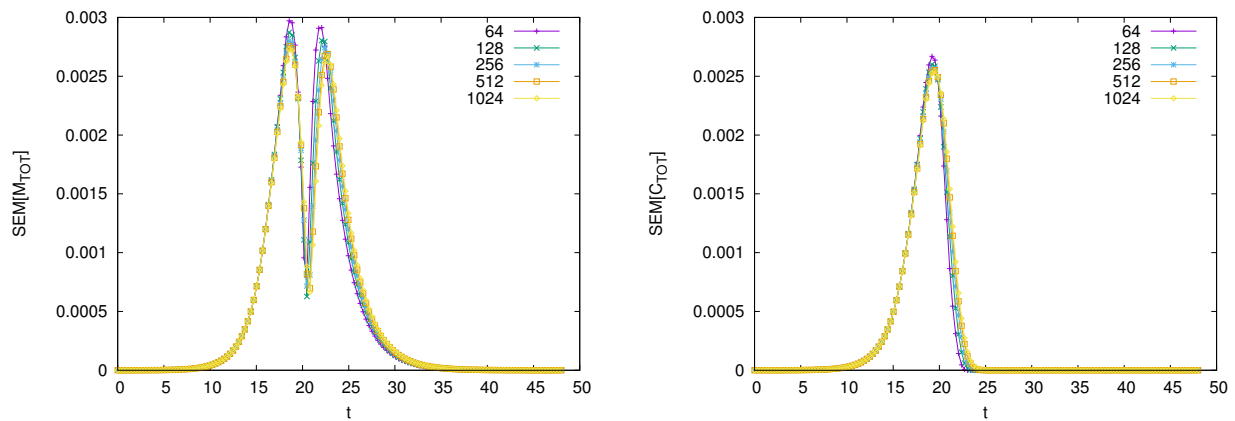


Figure 8. The total biomass SEM (left) and total substrate SEM (right) for several grid sizes obtained over 30 realizations. The spatial size of an attachment event is $\frac{1}{64} \times \frac{1}{64}$ of the domain.

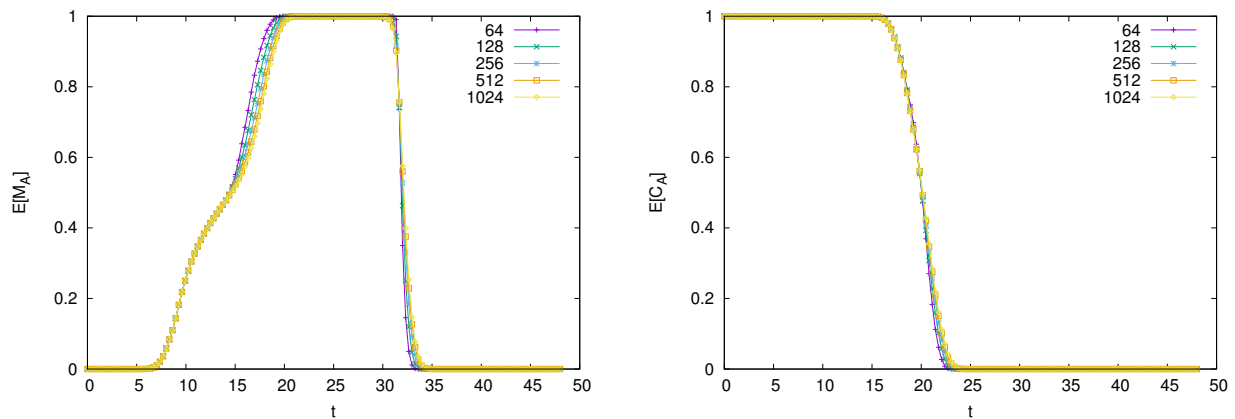


Figure 9. The time evolution of the expectation of occupancy of biomass (left) and substrate (right) for several grid sizes averaged over 30 realizations. The spatial size of an attachment event is $\frac{1}{64} \times \frac{1}{64}$ of the domain.

across the time series for finer grid resolutions. This is also reflected in the colonies standard error plots. We see variability in the number of colonies whenever they are above 0, which is to be expected based on the calculation of the statistics. The standard error plots for total mass and occupancy of biomass show an interesting two node behaviour. The two nodes correspond to the mass and occupancy of biomass increasing and decreasing whereas the valley is where the two quantities are maximized. The standard error plots for total mass and occupancy of carbon show a single peak which roughly corresponds to the maximum rate of decay in the carbon substrate.

From all the figures and tables presented in this section, we can see that as the grid resolution is refined, we have the expectation graphs and the standard error graphs approaching the associated graphs of 1024×1024 more closely. However, we do not see the graphs converging as closely as we may wish, but due to the computational cost of simulating 30 realizations on a 2048×2048 grid, we

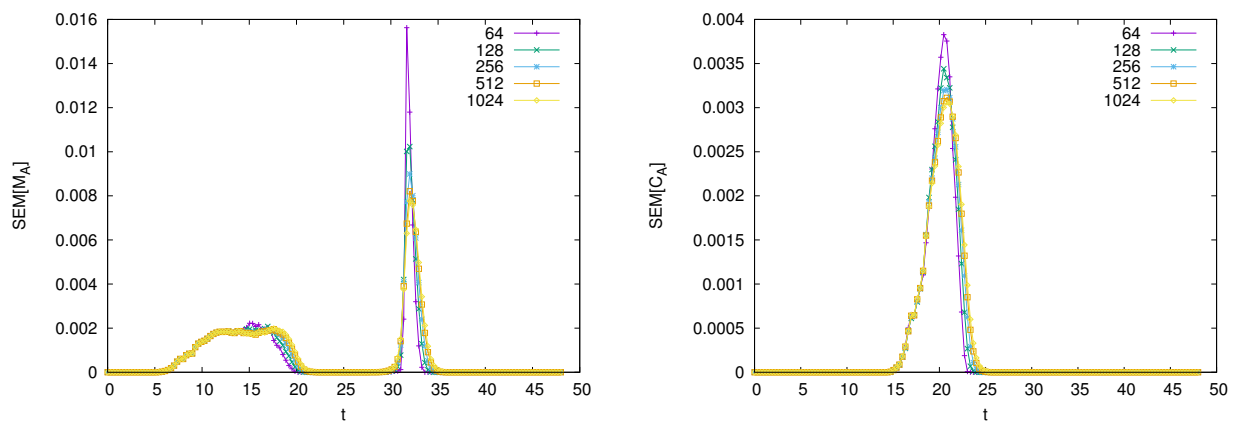


Figure 10. The occupancy of biomass SEM (left) and occupancy of substrate SEM (right) for several grid sizes obtained over 30 realizations. The spatial size of an attachment event is $\frac{1}{64} \times \frac{1}{64}$ of the domain.

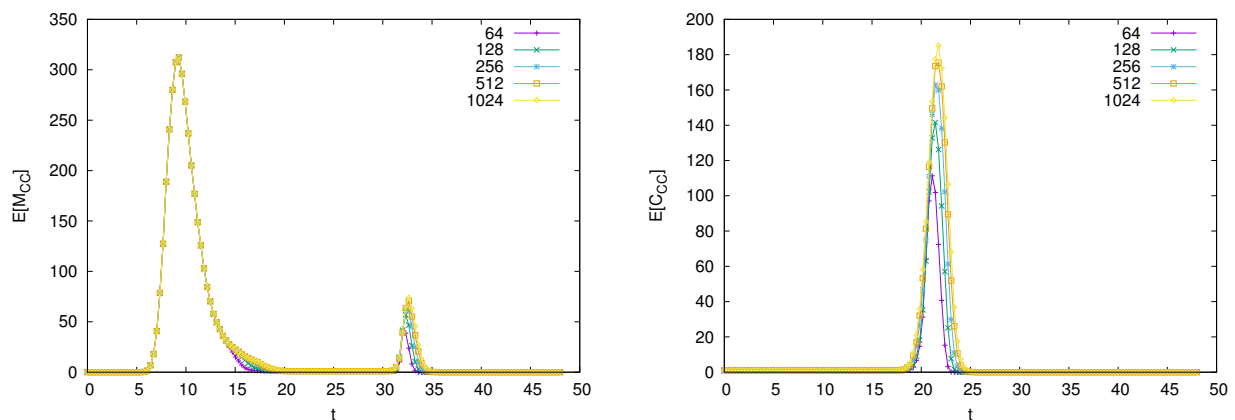


Figure 11. The time evolution of the expectation of the number of biomass colonies (left) and depleted substrate colonies (right) for several grid sizes averaged over 30 realizations. The spatial size of an attachment event is $\frac{1}{64} \times \frac{1}{64}$ of the domain.

cannot include any more grid refinement. Nevertheless, we still get relatively adequate convergence results, as seen in the figures and tables.

5.3. Attachment interval sensitivity analysis

In this section, we investigate the effects of the attachment interval on the three quantities of interest given in Section 5.2, namely the total biomass and carbon, the occupancy of biomass and carbon, and the number of biofilm colonies and fully depleted carbon colonies. In order to examine the effects of the attachment interval, we perform sensitivity analysis by varying the location of the attachment interval on the real line while keeping the length of the interval the same. For our purposes, we fix the length of the attachment interval as 0.00001 and translate the interval by $\xi > 0$. We assume the base

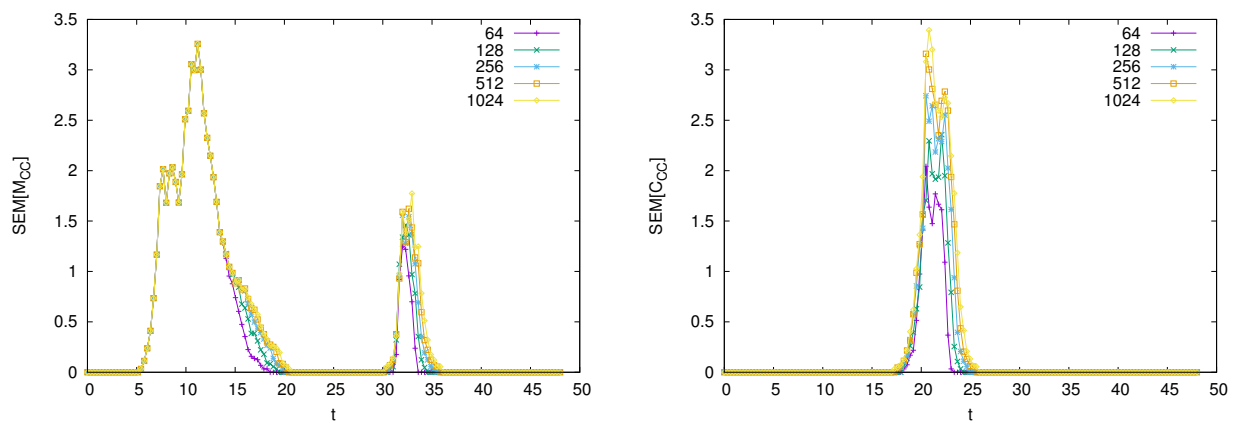


Figure 12. The number of biofilm colonies SEM (left) and depleted substrate colonies SEM (right) for several grid sizes obtained over 30 realizations. The spatial size of an attachment event is $\frac{1}{64} \times \frac{1}{64}$ of the domain.

interval is $[0.99999, 1.0]$ and generate our new interval as

$$[a, b] = [0.99999 + \xi, 1.0 + \xi]. \quad (5.9)$$

Since the initial value for each Wiener process is 0, we expect that as $[a, b]$ deviates from 0, the number of attachment events will decrease and the first attachment events will occur later due to the continuity of the Wiener process.

For all simulations, we set the initial conditions to $M_0 \equiv 0$ and $C_0 \equiv 1$ in order to simulate an initially clean domain. We use 30 realizations to generate the same test statistics as in Section 5.2. However, since we are not interested in investigating the convergence of the sample statistics, we only include a plot of the expectation time-series and include error bars at each point equal to \pm one standard deviation. We opt to use the standard deviation ($SEM(X) \sqrt{n}$) in this setting since the standard error in the sample mean is too small to detect at the current image size. The simulation results are summarized in Figures 13–15.

In all three figures, we can see that increasing ξ leads to slower initial changes in the three quantities of interest, smaller peak values and slower decay. This is caused by the longer time required for attachment events to occur and the lower number of attachment events when increasing ξ . We see from Figures 13 and 14 that the spatially integrated and occupancy plots follow each other quite closely in terms of shape, which is unsurprising due to the similarity of what they measure. The number of colonies graph, however, differs quite greatly from the other two. We see that the number of depleted carbon colonies increases rapidly shortly after the number of biomass colonies starts to decrease. This corresponds to the dominance of the $-\lambda M$ term in the biomass equation, which occurs once the carbon has become significantly depleted.

Although it is difficult to see in the plots of the total mass and occupancy, as we increase ξ , we see that the error bars in the mass and occupancy plots increase in length, which suggests there is more variability across the given realizations. This observation of increasing error bars is following what was found in [9]. Therefore, when more attachment events occur, the deviation in how much biomass is

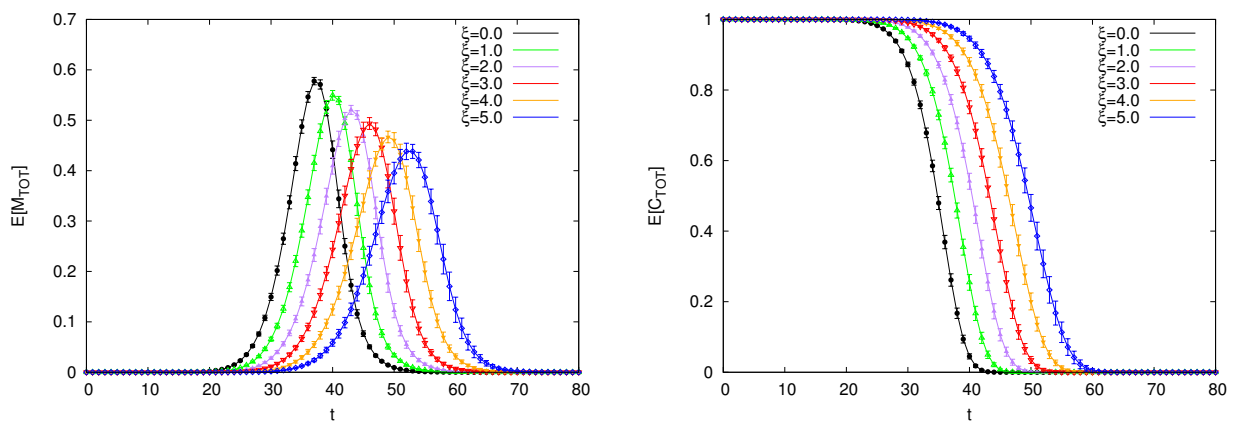


Figure 13. The time evolution of the expectation of total biomass (left) and substrate (right) for several attachment intervals averaged over 30 realizations. The vertical bars represent the mean \pm one standard deviation.

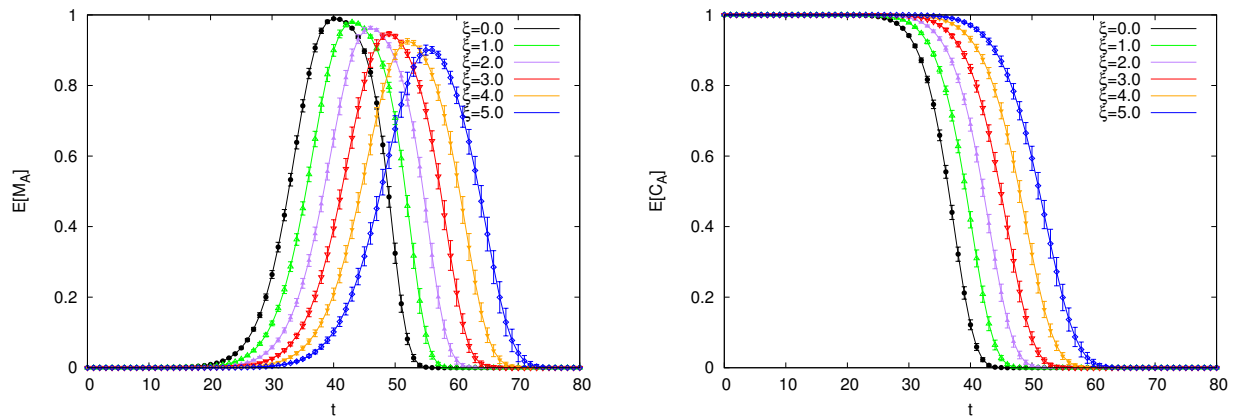


Figure 14. The time evolution of the expectation of occupancy of biomass (left) and substrate (right) for several attachment intervals averaged over 30 realizations. The vertical bars represent the mean \pm one standard deviation.

present and how much total biomass is occupying the domain decreases. However, when we look at the colonies' plots, we typically see more variability in the biomass plots when more attachment occurs. This indicates that the number of colonies can vary more greatly when we have a higher chance of attachment events occurring.

5.4. Stability of a travelling wave solution

Evidence of the existence of a 1D travelling wave solution for the deterministic models (2.1) and (2.2) has been given in [40] using a very similar numerical method to ours except they used the Implicit Euler method for the biomass equation. Furthermore, the existence of a travelling wave solution to the 1D PDE-ODE coupled system along with a linear stability result in the 2D case, has been recently shown analytically in [41]. However, there has been no analytical investigation into the stability of such

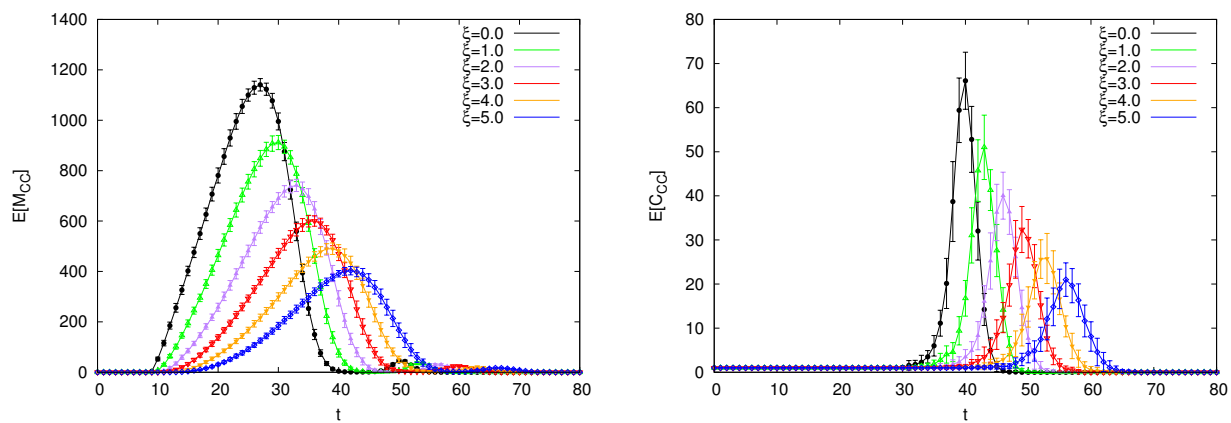


Figure 15. The time evolution of the expectation of the number of biomass colonies (left) and depleted substrate colonies (right) for several attachment intervals averaged over 30 realizations. The vertical bars represent the mean \pm one standard deviation.

a travelling wave solution under random perturbations. This is a challenging open problem and left for future research. So far, only few results have been obtained concerning the existence and stability of travelling waves for stochastic reaction-diffusion equations or PDE-ODE couplings, e.g. see [42–44] and the references therein. We are not aware of results for stochastic systems with degenerate diffusion.

This section is broken into two parts. First, we show that our numerical methods can accurately reproduce a travelling wave solution for the deterministic model (3.7) with $\mathcal{T}(\mathbf{M}, \mathbf{C})\mathbf{F}(\mathbf{W}_t) = 0$. We then use our random attachment model to introduce random perturbations in the system, which we use to provide evidence for the stability of the travelling wave. All simulations presented in this section use the parameter values in Table 1 unless otherwise stated.

5.4.1. Travelling wave solution of the deterministic model

We ran a simulation of (3.7) with $\mathcal{T}(\mathbf{M}, \mathbf{C})\mathbf{F}(\mathbf{W}_t) = 0$. The initial wave profile for both M and C was verified using the method described below. Figure 16 provides a summary of the simulation.

We note by the wave interface plot in Figure 17 that the 1D structure is preserved throughout the simulation. Therefore, we can stick to a given x -coordinate. To show that we have a 1D travelling wave solution, we need to show that the solution propagates at a fixed wave speed and does not change shape. We can do this graphically by horizontally translating the solution at different time points by $c(t_0 - t_n)$, where c is the wave speed, t_0 is the reference time point, and t_n are the times where that solution exists pre-translation. If the translated solutions coincide, then we can conclude that there is evidence of a travelling wave solution. Therefore, we need to first estimate the wave speed of our simulation. To do this, we calculate the location of interface of the biomass wave at the time points generated throughout the simulation. To calculate the location of the wave interface, we take the largest y -coordinate that has a biomass density larger than 10^{-2} . Once we have the location of the wave interface, we can estimate the wave speed by fitting a linear function through these data points. To fit a linear model to the data set, we use the built-in function, *fit*, from GNUPLOT. This fits the model $f(x) = cx + b$ to our data, where c is the estimate of the wave speed. The results are plotted in Figure 17.

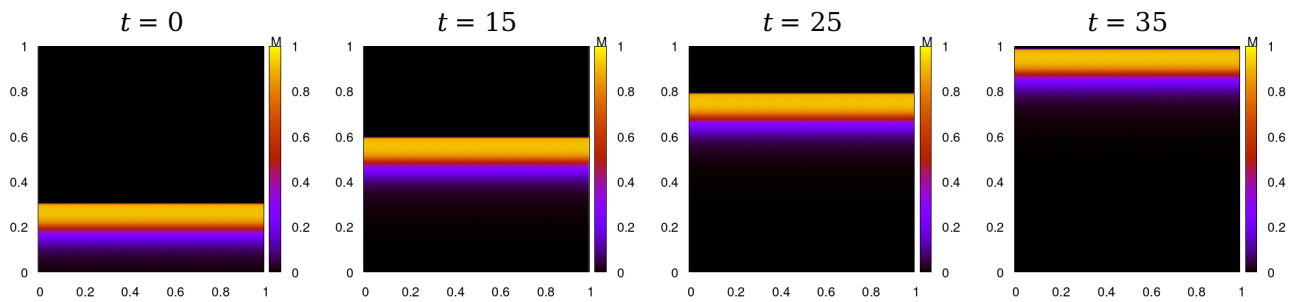


Figure 16. Simulation of the biomass density M from (2.11) with initially M and C given as the wave profile (see Figure 17).

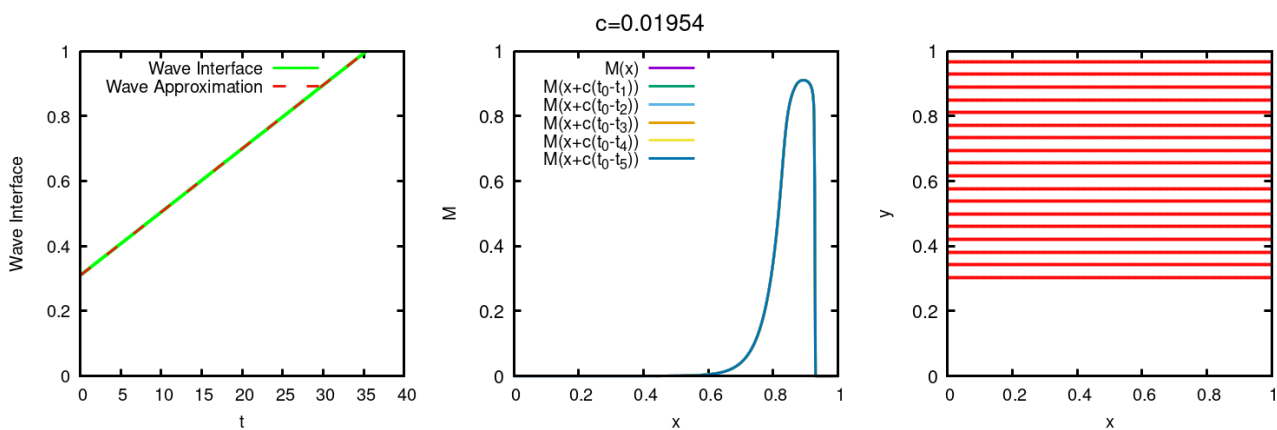


Figure 17. Travelling wave calculations for the simulation summarized in Figure 16. **Left:** The location of the wave peak as a function of t . The green line is the wave peak location taken from the simulation result. The dashed red line is the function $f(x) = cx + b$, with c as the wave speed, taken from the GNU PLOT fitting. **Middle:** The solutions of M in the form $M(x - ct)$ translated by a factor of $c(t_0 - t_n)$. The multiple time steps are translated on top of each other by horizontal translations of $c(t_0 - t_n)$ for each time step shown. The values of t_n are given by [25, 30, 35, 20, 15, 10]. **Right:** The wave interface for various time points $2n$ for $n = 0, \dots, 17$. The title shows the wave speed.

With our calculated wave speed, we can now investigate whether the shape of the wave is preserved. To do this we plot the solutions $M(x - c(t_0 - t_n))$, where $t_0 = 25.0$ is the reference time point, for which we translate the waves onto (see Figure 17). The values of t_n are the time points for the other solutions. By translating the solutions by a factor of $c(t_0 - t_n)$, we see that the solutions agree almost exactly, barring the thickness of the plotted lines. Therefore, we have sufficient evidence to suggest that the deterministic model, with our numerical method, can produce a travelling wave solution.

5.4.2. Random perturbations of a travelling wave solution

To start, we investigate one simulation with attachment to illustrate that the wave speed likely persists and the wave profile re-establishes. Due to the small spatial domain and time scale required for a travelling wave to establish, we start our simulation with the wave profile from Figure 16 both for the biomass density and the carbon concentration.

To simulate random perturbations of the wave, we apply attachment events only to the section of the domain where $y \in [256\Delta x, 341\Delta x]$, which corresponds to the y grid cells between $K/2$ and $(2K - 1)/3$, where $K = 512$. The impulse interval used in this next simulation is given by $[a, b] = [0.4999, 0.5]$. We chose this impulse interval and the specified grid cells as we get significant attachment, but the attachment sites do not form enough biomass to destroy the wave. Figure 18 illustrates this simulation. The colours chosen in the 2D colour map are different from the previous simulations to give a higher contrast for the attachment events.

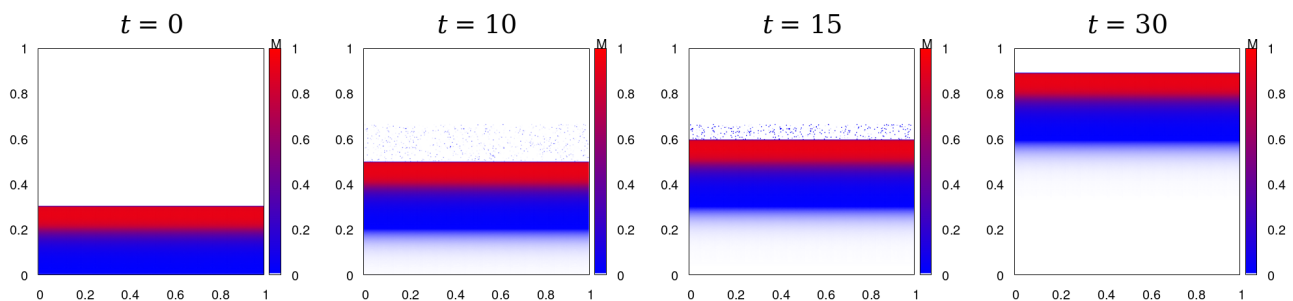


Figure 18. Simulation of (3.7) with initially M and C given by the wave profile and attachment only occurring in the region where $y \in [256\Delta x, 341\Delta x]$. The interval of attachment is $[0.4999, 0.5]$. We use a different color map to better show the attachment events in front of the wave.

From the snapshots given in Figure 18, we see that our travelling wave seems to persist through the attachment events. We now further investigate how the wave speed, interface, and profile changes after the attachment events.

We fit our linear model to all interface values after $t = 30$. Figure 19 illustrates the calculation of the wave speed. Knowing the wave speed, we can investigate whether we have the same wave profile across multiple time steps. We proceed as before. Using a reference time step $t_0 = 33.0$. Figure 19 illustrates the translated wave profiles. The translation of wave profiles gets slightly more complicated as the attachment does jump the wave profile forward, which means if we take the fitted line before and after the attachment sites, the y -intercepts are not the same. Therefore, if we translate the waves by the same factor $c(t_0 - t_n)$, our wave profiles will not necessarily coincide. We need to account for the difference in the y -intercepts. Taking the line of best fit over the interface points up to $t = 15$ gives us a y -intercept value for the pre-attachment wave of $b_1 = 0.3101$ while the y -intercept value post attachment is $b_2 = 0.3099$. If we translate the first set of points 33, 34, 35 which are post-attachment by a factor of $c(t_0 - t_n)$ while translating the pre-attachment time points 5, 10, 15 by a factor of $c(t_0 - t_n) + b_2 - b_1$ we get the plot given in Figure 19.

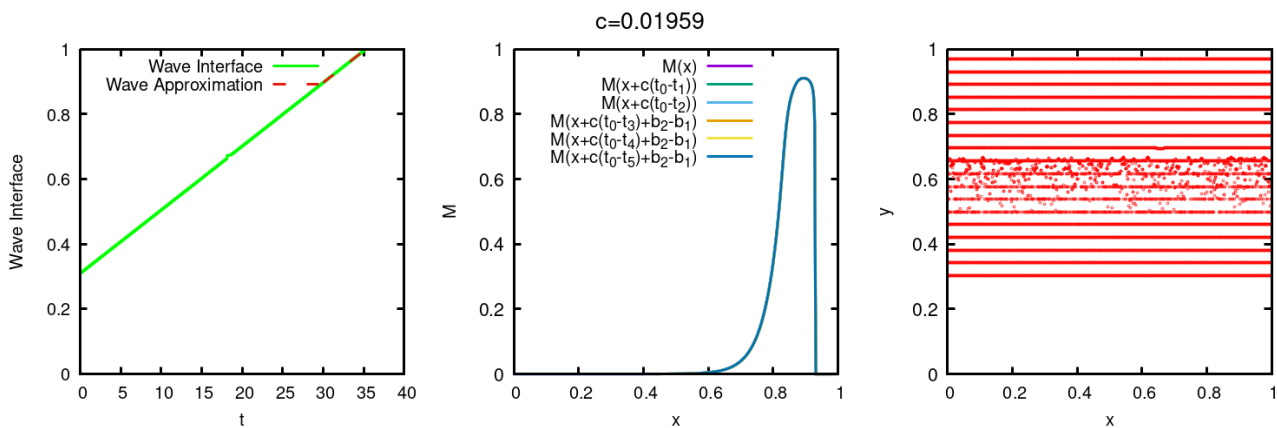


Figure 19. Travelling wave calculations for the simulation summarized in Figure 18. **Left:** The location of the wave peak as a function of t . The green line is the wave peak location taken from the simulation result. The dashed red line is the function $f(x) = cx + b$, with c as the wave speed, taken from the GNU PLOT fitting. **Middle:** The solutions of M in the form $M(x - ct)$ translated by a factor of $c(t_0 - t_n)$. The multiple time steps are translated on top of each other by horizontal translations of $c(t_0 - t_n)$ for post-attachment time points and $c(t_0 - t_n) + b_2 - b_1$ for pre-attachment time points. The values of t_n are given by [33, 34, 35, 5, 8, 11]. **Right:** The wave interface for various time points $2n$ for $n = 0, \dots, 17$. The title shows the wave speed.

As with our previous deterministic travelling wave simulations, we see that the profiles coincide completely across the time points, barring the thickness of the plotted curves, which suggests that the travelling wave is stable. We lastly provide a plot of the wave interface at various time points to illustrate that the interface returns to one dimension after re-establishing itself. The wave interface at various time points is given in Figure 19. The plot of the wave interface is calculated in the same manner as in the deterministic case. At a given time point t , we find the largest y -coordinate for each x -coordinate that has $M(t, x, y) > 10^{-2}$. Then we plot the (x, y) point. The distortions that appear in the region $y \in [0.5, 0.7]$ are caused by the attachment events distorting the wave interface. We see that after clearing the attachment region the wave interface re-establishes.

To give a more rigorous investigation into the stability of the travelling wave, we ran a hypothesis test to investigate whether or not the wave speed changes significantly post attachment. We ran 30 simulations with different initial seeds for the pseudo-random number generation used in the updates of the Wiener process.

With the calculation of the wave speed by linear model fitting, we get an error with the model and the associated parameters, along with the associated errors that arise from the numerical method and spatial discretization. We can also get differing wave speeds depending on which values we use for our linear fitting. If we start fitting at $t = 30$ for the simulation described in Figure 17, then we get a wave speed of $c = 0.01958$, which differs from the speed calculated from Figure 17. Therefore, we cannot expect more than 4 digits of accuracy in our wave speed calculations. Furthermore, the travelling wave

in an attachment simulation typically re-establishes itself at around $t = 30$. Therefore, for consistency sake, we choose $c = 0.01958$ as our true wave speed[‡] and calculate all post-attachment wave speeds for time points after $t = 30$ and round to the nearest 10^{-5} . The wave speeds are given in Table 3.

Table 3. Post attachment wave speed values calculated over 30 realizations. The wave speeds were calculated over all interface points after $t = 30$ in the simulations.

Seed	Wave Speed	Seed	Wave Speed	Seed	Wave Speed
1	0.01959	11	0.01958	21	0.01958
2	0.01959	12	0.01957	22	0.01958
3	0.01957	13	0.01959	23	0.01957
4	0.01959	14	0.01958	24	0.01959
5	0.01958	15	0.01958	25	0.01957
6	0.01959	16	0.01958	26	0.01959
7	0.01959	17	0.01957	27	0.01957
8	0.01959	18	0.01958	28	0.01959
9	0.01957	19	0.01958	29	0.01956
10	0.01959	20	0.01959	30	0.01957

With these data, we can conduct a hypothesis test to determine if there is a significant statistical difference between computed post-attachment wave speeds and the deterministic wave speed. We assert the null hypothesis that the true mean wave speed post attachment is the same as the true wave speed, and use a two-sided alternative hypothesis. To generate the statistics, we use the *OneSampleTTest* function from the Hypothesis Tests package in Julia. Given our computed post attachment wave speeds c_i , we get the following statistics:

$$\bar{c} = 0.0195813, \quad SEM(\bar{c}) \approx 1.64235 \times 10^{-6}, \quad t_{\bar{c}} \approx 0.811844, \quad p \approx 0.4235. \quad (5.10)$$

Given our large p -value of 0.4235, we fail to reject the null hypothesis that the true mean wave speeds are the same post attachment. Constructing a 95% confidence interval about the sample mean gives (0.019578, 0.019585). Given our previous discussion on the associated errors from the numerical methods and the linear fitting, we conclude that our calculated post attachment wave speeds are within the margin of error of the deterministic wave speed. Thus, concluding that we have significant evidence in favour of a stable travelling wave solution.

6. Discussion

The main objective of this work was to develop a new mathematical modelling framework for discrete stochastic attachment that improved on the drawbacks of the previous attempts in [9, 16]. Namely, a lack of a rigorous mathematical framework in [16] and expensive simulation runtime in [9]. Through applying our framework to the PDE-ODE coupled system for cellulolytic biofilm growth we gained more theoretical insight into our model while also greatly improving the simulation runtime when compared to the model presented in [9].

[‡]The wave speed of the deterministic simulation calculated for interface values after $t = 30$ and rounded to the nearest 10^{-5} .

The nature of stochastic cellular attachment, along with the lack of experimental data, makes it difficult to propose a cellular attachment framework based on underlying mechanisms in the system. This makes the stochastic approach used here adequate for now. However, if future experiments and analyses can provide a mechanism for cellular attachment in a fully deterministic way, the resulting model will likely capture the system behaviour in a more concise manner.

The use of the RODE framework lent itself better to capturing attachment events when compared to the Itô SDE framework used in [9]. Our framework omits the addition of an extra NK DEs used to model the attachment signal; we were able to lump that into a stochastic process for each grid cell. This modification simplifies the mathematical theory we can rely on since RODEs are path-wise ODEs. We were able to show the global existence and uniqueness of a regularized version of our model and the long-term behaviour of solutions, which was unattainable with the SDE framework.

The elimination of the extra NK DEs from the model in [9] does come with some notable disadvantages, mostly related to the fitting of the model. The attachment factor used in [9] had additional parameters and depended on state variables, a mechanism that is not present in our formulation. In our model, if the biomass was close to unity or the substrate was close to zero, the Wiener process was still simulated and thus, could still cause an attachment event to take place. Now, since the biomass and substrate would hinder the amount of biomass attached (by a factor of $g(M)h(C)$), we would not have much biomass attached; however, there would still be some. However, with the model from [9], the attachment factor was also controlled by the state variables in a similar manner as the attachment term in our model. Therefore, when the biomass and carbon were inhibiting attachment, they were also inhibiting the attachment factor, which for most realizations meant no attachment would occur instead of the possibility for attachment events.

For simplicity, we used the Wiener process, as we can simply use a random number generator to simulate it. Moreover, the qualitative behaviour of the model is as desired. We tried other stochastic processes, including the Ornstein-Uhlenbeck process and found that the qualitative behaviour of the model did not change; however, the simulation time did increase. It is an interesting question to see how different stochastic processes can affect the model behaviour. However, this is outside the scope of our research.

The use of an implicit time integrator for our model allowed us to drastically improve simulation times when compared to the explicit solver used in [9] (by a factor of over 300%). Unfortunately, some of the traditional disadvantages of an implicit method became apparent in our situation. The major issue arises with larger time steps; we can lose out on some of the attachment events if the computed values of the Wiener process bracket the impulse interval without being in the interval, which means we introduce more errors. The lengths of the attachment intervals used in Section 5.1 are 0.0001 and 0.00001. When we simulate the Wiener process, we use the formula $W_{k+1} = W_k + \sqrt{\Delta t} \Delta W_k$ where ΔW_k is a standard normal random variable. The probability that $|\Delta W_k| > 1$ is roughly 31.8%. Therefore, to avoid missing attachment events 68.2% of the time, we need Δt to be either 10^{-8} or 10^{-10} , which is way too low for an implicit time step and is much closer to the time steps used in [9]. This is the main reason why changing the impulse interval in Section 5.1 lead to significantly more attachment events since the Wiener processes more likely attained a value within the interval when the interval is larger. To mitigate this issue, we could use a different stochastic process, preferably one with less variance, so that the increments are scaled by a smaller number, which would lower the likelihood of missing attachment events. We could have also used a finer step size to simulate the Wiener process.

Meaning, if Δt is our implicit time step, then we could use the time step $\Delta t/N$ to simulate the Wiener process. This does come with some challenges though. For example, if we have an attachment event occurring on some fraction of Δt , then it is unclear how much should be attached at the Δt time step. Moreover, we need to decide how to handle the case where a Wiener process enters and exits the attachment interval more than once throughout a Δt time step. Lastly, in order to ensure that we have an appropriate number of attachment events we will need to shrink the size of our attachment interval. However, shrinking the size of the interval implies a higher chance of missing attachment events. For these reasons, we simulated our Wiener processes using the Δt time step as it gave us the desired qualitative behaviour.

From a practical standpoint, the large time step does not affect our model in a meaningful way. Since we have no experimental data to calibrate our model to, we cannot currently use our model to predict quantitative behaviour. Furthermore, spatio-temporally resolved experimental studies are very difficult and the current microscopy is not able to give the spatial and dynamic resolution that would be needed for a direct calibration. However, if more experimental data were available or in cases where we wished to use our framework in a more sophisticated system, we may have concerns with missing attachment events. These concerns also manifest themselves when we consider the reduced control we have over the rate of attachment events, as we have less chance of properly calibrating our model or properly capturing system behaviour in complex systems. A more complex system where these concerns will likely matter is any system where we have bulk liquid surrounding the biofilm such that the availability of free cells is determined by this liquid.

Since we investigated the effects of the attachment interval, knowing that we likely missed attachment events, we provide some validation to our methods here. We chose to fix the length of the attachment interval and vary its position on the real line as we conjectured that preserving the length of the attachment interval would lead to roughly the same proportion of missed attachment events. So, here we investigate that claim. We took twelve different attachment intervals of the form

$$[a, b] = [0.49999 + \xi, 0.5 + \xi]$$

with an integer ξ in the range 0–11 and simulated 512×512 Wiener process over 200 time units with a time step of 10^{-2} . We counted the number of attachment events along with the number of missed attachment events[§] and took the ratio of missed attachment events over counted attachment events. The data are presented in Table 4. As we can see, we roughly miss 8000 attachment events for every 1 attachment event we count. However, across all twelve ξ values, we see that the proportion of missed events is roughly equal, which suggests that our investigation into qualitative effects of the attachment interval is justified.

Table 4. Ratio of missed attachment events over counted attachment events for 12 different impulse intervals.

ξ	Ratio	ξ	Ratio	ξ	Ratio	ξ	Ratio
0.0	7915.35	1.0	7946.69	2.0	7903.99	3.0	7921.85
4.0	8280.92	5.0	8030.46	6.0	8085.27	7.0	7709.58
8.0	8176.43	9.0	8178.81	10.0	8177.22	11.0	7821.86

[§]Missed attachment events are when the current and next value of the Wiener process bracket the impulse interval.

Throughout Section 5 we often ran 30 simulations with the same parameter values to generate different sample statistics or to run hypothesis tests. As outlined in that chapter, we chose 30 simulations as that is the unwritten rule to apply the Central Limit Theorem. However, the Central Limit Theorem typically applies when we are interested in hypothesis testing, but we chose 30 simulations for consistency's sake. Here we provide further justification for the use of 30 realizations, which is summarized in Figure 20. To generate these plots we ran 100 simulations with the parameter values in Table 1 with $[a, b] = [0.99999, 1.0]$ and calculated the 1-norm of the standard deviation time-series for each quantity of interest in Chapter 5. The simulations were run until $t = 70$. We see that the standard deviation does not change drastically after 30 samples, supporting our decision.

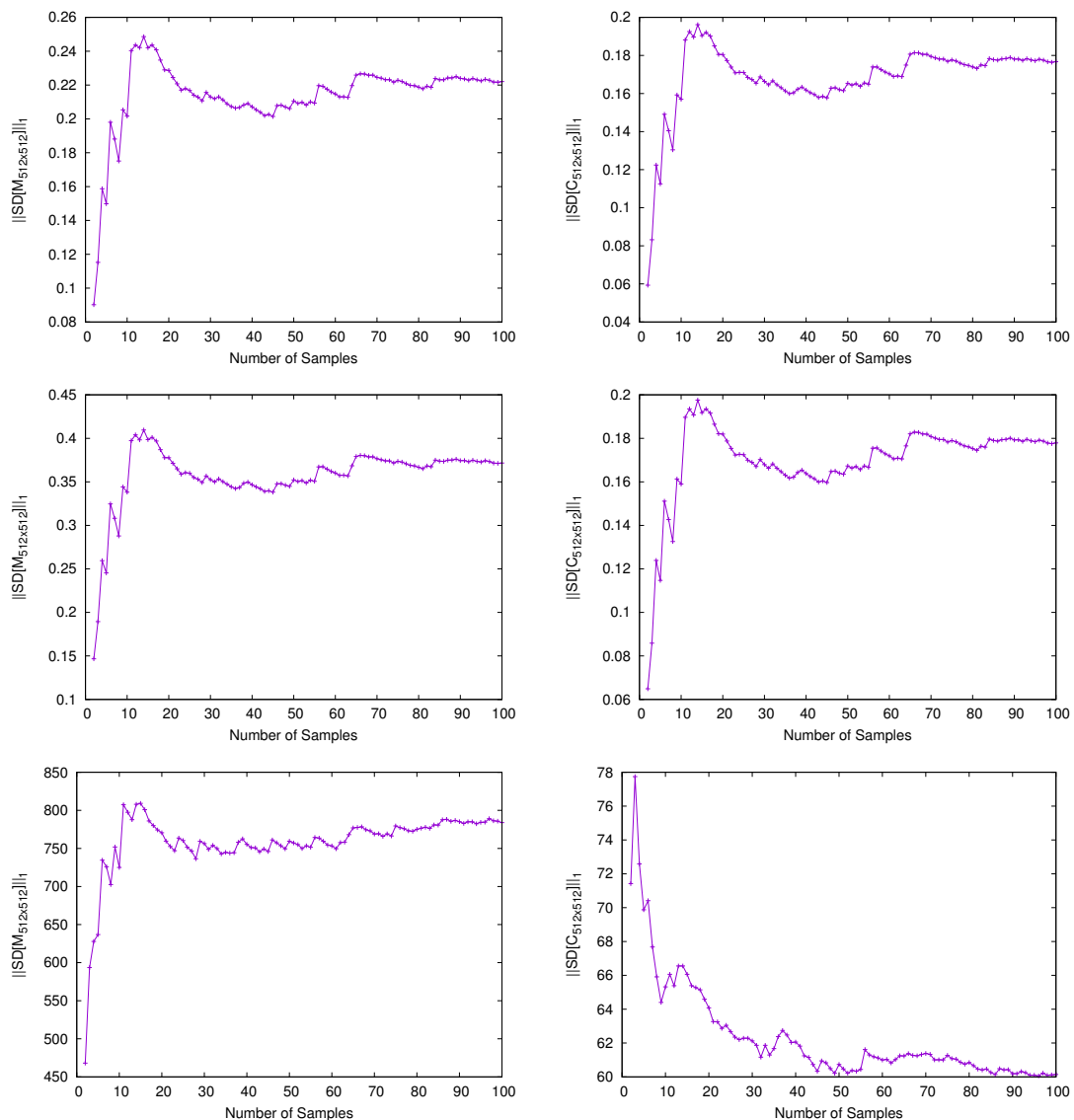


Figure 20. The 1-norm of the standard deviation times series using standard deviations calculated with different sample counts **Top:** Total Mass, **Middle:** Occupancy, and **Bottom:** Colonies. The x -axis represents the number of samples used to calculate the standard deviation. The biomass curve is on the left and the carbon curve is on the right.

In our work, the deterministic model is the spatial discretized PDE-ODE coupled system (2.11). We could argue that our deterministic model is the PDE-ODE coupled systems (2.1) and (2.2). The reason why we work with the discretized version is the discrete nature of attachment. PDEs work under the assumption of infinitesimals, which does not lend well to the spatially discrete attachment. However, if we fixed the size of an attachment event, we likely could generalize our model to a PDE via the use of non-local effects. We could take the original PDE-ODE coupled system and add some linear combination of convolutions to add attachment to all locations in a grid cell via the influence of a Wiener process. This construction was outside the scope of this work; however, it would be very interesting to develop and analyze such a model and see if our spatially discrete system (3.7) converges as $\Delta x \rightarrow 0$.

7. Conclusions

Discretized deterministic continuous biofilm models can be extended into RODEs in order to capture discrete stochastic attachment events via an impulse function with a standard stochastic process as input. We see that this formulation preserves the non-negativity of the system, lends itself well to extending results on the deterministic model, and improves simulation runtime greatly when compared to earlier formulations. Our attachment model (3.7) along with the proposed numerical methods, can also be used to investigate the model behaviour of the underlying deterministic model. Namely, we gave evidence to suggest that there exists a stable 1D travelling wave solution to the PDE-ODE coupled systems (2.1) and (2.2).

The model developed in this paper is for the particular case of cellulosic biofilms. We think this modelling framework could be applied to other two or three dimensional continuum models, including traditional biofilm models where nutrients are supplied via diffusion from the aqueous phase.

Acknowledgments (All sources of funding of the study must be disclosed)

This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) CGS-M awarded to JMH, and by an NSERC Discovery Grant (RGPIN-2019-05003) and NSERC Research Tools and Infrastructure Grant (RTI-2016-00080) awarded to HJE. It was supported by the *Nederlands Organisatie voor Wetenschappelijke Onderzoek* (NWO) with grant OCENW.KLEIN.358, awarded to SS.

Conflict of interest

No conflicts are declared.

References

1. D. López, H. Vlamakis, R. Kolter, Biofilms, *CSH Perspect. Biol.*, **13** (2010), 1–11. <https://doi.org/10.1101/cshperspect.a000398>
2. L. Hall-Stoodley, J. W. Costerton, P. Stoodley, Bacterial biofilms: Form the natural environment to infectious diseases, *Nat. Rev. Microbiol.*, **2** (2004), 95–108. <https://doi.org/10.1038/nrmicro821>

3. R. M. Donlan, Biofilms: Microbial life on surfaces, *Emerg. Infect. Dis.*, **8** (2002), 881–890. <https://doi.org/10.3201/eid0809.020063>
4. O. Wanner, H. Eberl, E. Morgenroth, D. R. Noguera, C. Picioreanu, B. Rittmann, et al., *Math. model. biofilms*, IWA Publishing, 2006.
5. M. C. Van Loosdrecht, J. J. Heijnen, H. Eberl, J. Kreft, C. Picioreanu, Mathematical modelling of biofilm structures, *Antonie van Leeuwenhoek*, **81** (2002), 245–256. <https://doi.org/10.1023/a:1020527020464>
6. Y. H. An, R. J. Friedman, Concise review of mechanisms of bacterial adhesion to biomaterial surfaces, *J. Biomed. Mater. Res.*, **43** (1998), 338–348. [https://doi.org/10.1002/\(sici\)1097-4636\(199823\)43:3<338::aid-jbm16>3.0.co;2-b](https://doi.org/10.1002/(sici)1097-4636(199823)43:3<338::aid-jbm16>3.0.co;2-b)
7. A. Dumitrache, G. Wolfaardt, G. Allen, S. N. Liss, L. R. Lynd, Form and function of *Clostridium thermocellum* biofilms, *Appl. Environ. Microbiol.*, **79** (2013), 231–239. <https://doi.org/10.1128/AEM.02563-12>
8. Z. W. Wang, S. H. Lee, J. G. Elkins, J. L. Morrell-Falvey, Spatial and temporal dynamics of cellulose degradation and biofilm formation by *Caldicellulosiruptor obsidianis* and *Clostridium thermocellum*, *AMB Express*, **1** (2011), 1–10. <https://doi.org/10.1186/2191-0855-1-30>
9. Y. Rohanizadegan, S. Sonner, H. J. Eberl, Discrete attachment to a cellulolytic biofilm modeled by an Itô stochastic differential equation, *Math. Biosci. Eng.*, **17** (2020), 2236–2271. <https://doi.org/10.3934/mbe.2020119>
10. B. K. Øksendal, *Stochastic differential equations: an introduction with applications*, Springer, 2013.
11. A. Carroll, C. Somerville, Cellulosic biofuels, *Annu. Rev. Plant. Biol.*, **60** (2009), 165–182. <https://doi.org/10.1146/annurev.arplant.043008.092125>
12. M. H. Langholtz, B. J. Stokes, L. M. Eaton, 2016 billion-ton report: Advancing domestic resources for a thriving bioeconomy, *Oak Ridge National Lab.*, 2016.
13. J. G. Linger, A. Darzins, *Consolidated Bioprocessing*, Springer, 2013.
14. B. G. Schuster, M. S. Chinn, Consolidated bioprocessing of lignocellulosic feedstocks for ethanol fuel production, *Bioenergy Res.*, **6** (2012), 416–435. <https://doi.org/10.1007/s12155-012-9278-z>
15. V. Mbaneme-Smith, M. S. Chinn, Consolidated bioprocessing for biofuel production: Recent advances, *Energy Emission Control Technol.*, **3** (2015), 23–44. <https://doi.org/10.2147/EECT.S63000>
16. H. J. Eberl, E. M. Jalbert, A. Dumitrache, G. M. Wolfaardt, A spatially explicit model of inverse colony formation of cellulolytic biofilms, *Biochem. Eng. J.*, **122** (2017), 141–151. <https://doi.org/10.1016/j.bej.2017.03.007>
17. L. N. Trefethen, *Finite difference and spectral methods for ordinary and partial differential equations*, Cornell University Department of Computer Science and Center for Applied Mathematics, 1996.
18. B. D’Acunto, V. Luongo, M. R. Mattei, Free boundary approach to modelling multispecies biofilms, *Ric. Mat.*, **70** (2020), 267–284. <https://doi.org/10.1007/s11587-020-00523-7>

19. A. Mašić, H. J. Eberl, Persistence in a single species CSTR model with suspended flocs and wall attached biofilms, *Bull. Math. Biol.*, **74** (2012), 1001–1026. <https://doi.org/10.1007/s11538-011-9707-8>
20. H. J. Gaebler, H. J. Eberl, A simple model of biofilm growth in a porous medium that accounts for detachment and attachment of suspended biomass and their contribution to substrate degradation, *European J. Appl. Math.*, **29** (2018), 1110–1140. <https://doi.org/10.1017/S0956792518000189>
21. H. J. Gaebler, H. J. Eberl, Thermodynamic inhibition in chemostat models: With an application to bioreduction of uranium, *Bull. Math. Biol.*, **82** (2020), 1–25. <https://doi.org/10.1007/s11538-020-00758-3>
22. H. J. Eberl, D. F. Parker, M. C. Van Loosdrecht, A new deterministic spatio-temporal continuum model for biofilm development, *J. Theor. Med.*, **3** (2001), 161–175. <https://doi.org/https://doi.org/10.1080/10273660108833072>
23. H. J. Eberl, L. Demaret, A finite difference scheme for a degenerated diffusion equation arising in microbial ecology, *Electron. J. Differential Equations*, **15** (2007), 77–95.
24. J. Monod, The growth of bacterial cultures, *Annu. Rev. Microbiol.*, **3** (1949), 371–394.
25. A. Dumitrache, H. J. Eberl, D. G. Allen, G. M. Wolfaardt, Mathematical modeling to validate on-line CO₂ measurements as a metric for cellulolytic biofilm activity in continuous-flow bioreactors, *Biochem. Eng. J.*, **101** (2015), 55–67. <https://doi.org/10.1016/j.bej.2015.04.022>
26. M. Ghasemi, H. J. Eberl, Time adaptive numerical solution of a highly degenerate diffusion-reaction biofilm model based on regularisation, *J. Sci. Comput.*, **74** (2018), 1060–1090. <https://doi.org/10.1007/s10915-017-0483-y>
27. X. Han, P. E. Kloeden, *Random ordinary differential equations and their numerical solution*, 1st edition, Springer, 2017.
28. M. Lefebvre, *Applied Stochastic Processes*, Springer, 2007.
29. A. Dumitrache, G. M. Wolfaardt, D. G. Allen, D. G. Liss, L. R. Lynd, Tracking the cellulolytic activity of *Clostridium thermocellum* biofilms, *Biotechnol. Biofuels*, **6** (2013), 1–15. <https://doi.org/10.1186/1754-6834-6-175>
30. M. A. Efendiev, S. V. Zelik, H. J. Eberl, Existence and longtime behavior of a biofilm model, *Commun. Pure Appl. Anal.*, **8** (2009), 509–531. <https://doi.org/10.3934/cpaa.2009.8.509>
31. W. Walter *Ordinary differential equations*, 1st edition, Springer, 1998.
32. Y. Asai, E. Hermmann, P. E. Kloeden, Stiff integration of stiff random ordinary differential equations, *Stoch. Anal. Appl.*, **31** (2013), 293–313.
33. Y. Asai, P. E. Kloeden, Numerical schemes for random ODEs via stochastic differential equations, *Commun. Appl. Anal.*, **17** (2013), 511–528.
34. L. Shampine, S. Thompson, J. Kierzenka, G. Byrne, Non-negative solutions of ODEs, *Appl. Math. Comput.*, **170** (2005), 556–569. <https://doi.org/10.1016/j.amc.2004.12.011>
35. L. N. Trefethen, D. Bau, *Numerical Linear Algebra*, SIAM Society for Industrial and Applied Mathematics, 2000.

36. J. Hughes, *A mathematical model of discrete attachment to a cellulolytic biofilm using random DEs*, M. Sc. Thesis, University of Guelph, 2021. <https://hdl.handle.net/10214/26321>
37. J. Bezanson, A. Edelman, S. Karpinski, V. B. Shah, Julia: A fresh approach to numerical computing, *SIAM Rev.*, **59** (2017), 65–98. <https://doi.org/10.1137/141000671>
38. R. V. Hogg, E. A. Tanis, D. L. Zimmermanm *Probability and statistical inference*, 9th edition, Pearson, 2020.
39. W. Burger, M. Burge, *Principles of digital image processing: core algorithms*, Springer-Verlag, 2009.
40. E. Jalbert, *Comparison of a semi-implicit and a fully-implicit time integration method for a highly degenerate diffusion-reaction equations coupled with an ordinary differential equations*, M. Sc. Thesis, University of Guelph, 2016. <http://hdl.handle.net/10214/9448>
41. K. Mitra, J. M. Hughes, S. Sonner, H. J. Eberl, J. D. Dockery, Travelling waves in a PDE-ODE coupled system with nonlinear diffusion, preprint, arXiv:2202.07748.
42. K. Eichinger, M. V. Gnann, C. Kuehn, Multiscale analysis for traveling-pulse solutions to the stochastic FitzHugh-Nagumo equations, preprint, arXiv:2002.07234.
43. C. H. S. Hamster, H. J. Hupkes, Stability of traveling waves on exponentially long timescales in stochastic reaction-diffusion equations, *SIAM J. Appl. Dyn. Syst.*, **19** (2020), 2469–2499. <https://doi.org/10.1137/20M1323539>
44. C. H. S. Hamster, H. J. Hupkes, A general framework for stochastic traveling waves and patterns, with application to neural field equations, *SIAM J. Appl. Dyn. Syst.*, **15** (2016), 195–234. <https://doi.org/10.1137/15M102856X>



AIMS Press

©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)