Universitat
de les Illes Balears

Modelling Quorum Sensing Mechanisms in Bacterial Populations

AUTHOR: Victor Buendía Ruiz-Azuaga

Master**'s** Thesis

Master's degree in Physics of Complex Systems

at the

UNIVERSITAT DE LES ILLES BALEARS

Academic year 2016-2017

Date 14/07/2017

*UIB Master's Thesis Su*pervisor: Manuel Alberto Matías Muriel

UIB Master's Thesis Co-Supervisor: Ricardo Martínez García

Acknowledgements

Este trabajo está dedicado a todos los que me habéis ayudado a realizarlo, directa o indirectamente.

De forma directa, a mis directores. Primero, por su labor profesional de guiarme en el desarrollo del trabajo, pero también todo el apoyo personal a lo largo del año. Manuel y Ricardo me han ayudado en todos los sentidos, no sólo a la realización del Trabajo de Fin de Máster, sino también en otros aspectos más amigables y menos relacionados con la física. Asimismo, de manera directa, es menester agradecer al personal del IFISC las clases del máster, que han sido realmente productivas y útiles para mi futuro. Salgo del centro con la fuerte convicción de que estoy preparado para la siguiente fase de mi formación.

Por otro lado, las contribuciones de forma indirecta al trabajo han sido más numerosas. Para empezar, a mi familia, que siempre ha estado detrás apoyándome en lo que he querido hacer, incluso cuando les preocupaba que llegara o no a conseguirlo. Me han dado la oportunidad de elegir lo que me gustaba y gracias a ello ahora estoy aquí. En especial, me gustaría mencionar a mi hermana, que este año ha tenido un curso, con total seguridad, más difícil que el mío, y que me ha dado casi todos los días una buena conversación por teléfono. Ana, sé que no quieres reconocerlo, pero no quieres colgar porque si lo hicieras tendrías que ir a estudiar... Y todos lo sabemos.

A Adela y a Antonio, que sé que si pudieran hasta me habrían adoptado, y han estado pendientes de mí durante toda mi estancia aquí.

También para mis compañeros de máster, por la piña que se ha formado desde el primer momento y que ha durado hasta el último. Aunque podría decir que esto ha sido beneficioso académicamente, a decir verdad, las distracciones y ratos de risas en la habitación de «zulo» de máster, las quedadas semanales y fiestas han sido, con seguridad, de muchísima más utilidad. Gracias a ellos, un año que podría haber sido excepcionalmente aburrido y gris ha resultado ser una de las mejores experiencias durante el curso de máster. Por tanto, me parece importante agradecerles a todos ellos el buen año que hemos pasado juntos. Sí, incluso a aquellos que os habéis vendido al Big Data, a vosotros os quiero también.

Y por último, pero por supuesto no menos importante, gracias a Elvira, por su incondicional apoyo desde el primer momento que pensé en venir a estudiar a Mallorca, por su paciencia, por el esfuerzo que ha hecho por venir a verme, por decorar mi cuarto, por las largas conversaciones de teléfono por las noches... Por todo lo que has hecho por mí este año. Elvira, sin ti, todo esto hubiera sido muy diferente.

Gracias a todos vosotros, porque habéis hecho que el año en Mallorca haya sido excelente, y haya merecido la pena por encima de cualquier otra posibilidad de las miles que tenía para escoger un máster. Porque me voy con la sensación de que esta ha sido la decisión correcta. Este trabajo es para todos vosotros, y ojalá todos los años puedan ser tan buenos o mejor que este.

Abstract

Quorum sensing is a cell-to-cell communication mechanism that allows bacteria to trigger collective behaviours at the population scale. In this work, we present and analyse a model for one of the quorum sensing pathways in the bacterium *Pseudomonas aeruginosa*. This particular pathway controls the production of elastase, which is a «public good». Public goods are metabolic costly exoproducts for the individual, but, at high enough densities, they provide an overall benefit for the whole population. The model we present is an individual-based stochastic model, that also accounts for the spatial dimensions of the system. The objective of the model is to describe the growth of wild type and mutant strains, in single and mixed populations, using an implicit representation of the bacterial exoproducts used for the interactions. The individual based model allows us to characterize and understand the growth of a wild type strain that regulates the production of public goods via a quorum sensing mechanism, as well as to address classical questions about cooperation and defection in mixed cultures: can a non-producing mutant bacteria exploit the public good benefits? Moreover, is it possible to take advantage of the communication channel to outcompete the wild type strain? Which is the role of the spatial structure in this problem, and what is the relation with the diffusion of the exoproducts? Our model is able to give qualitative answers to these questions. We will compare these answers to some of the experiments and theoretical advances in the literature.

Contents

1. Introduction

1.1 The Quorum Sensing Mechanism.

1.1.1 Quorum Sensing: communication between bacteria.

Quorum sensing (QS) is a cell-to-cell communication mechanism that allows a single individual to track the density of neighbours and trigger collective behaviours at the population level. [1]. The quorum sensing mechanism is one of the most paradigmatic examples of interaction and communication between cells, and it is very common in bacteria [2]. Bacteria produce signal molecules that diffuse into the medium [1–3]. The concentration of signal in the medium can be detected by neighbouring cells and trigger several behaviours [1, 2]. This mechanism is therefore a way to trigger collective behaviours in the population. For example, quorum sensing regulates the expression of genes related to bioluminescence, motility, or public good production [1, 2]. A single organism can have different QS «channels» or pathways, regulated by different signal molecules [1,2]. However, a single pathway can regulate the production of several exoproducts.

In this work we focus in bacteria in which QS regulates the production of public goods. Public goods (PGs) are exoproducts released by bacteria that are beneficial for the whole population. The release of PGs by cells may look counterintuitive, since the PG molecule is costly to produce, and for a single cell the costs of production may be higher than the benefit provided [4], so the individual that releases PGs usually obtains a small share of the benefit generated, while the neighbours will obtain most of the benefit [4]. This means that the production may be exploited by non-producing cheaters, which would get the benefit without paying the costs of production [4, 5]. At first glance, the PG production behaviour may look altruistic. However, it is important to differentiate between a cooperative and a purely altruistic behaviour. In the case of the cooperative behaviour, the bacteria producing PGs will receive a benefit later, and it may even overcome the cost of producing the public goods [4,5]. The altruistic one, in contrast, does not imply any direct benefit for the producer individuals.

It is also important to stress that in the literature there is an abuse of the term «public good» for «external good» [5]. An external good should be labeled «public» only when it is shared equally among all the cells in the population [5]. In order to avoid an invasion by the non-producer, one of the possible strategies is precisely to limit the diffusion of the external good, making it «private» so it is only available to the producers [4, 5]. The self-organization of cells in space plays an important role in this kind of strategies, leading to clustered configurations of the population [4, 6]. In our model, the external good is shared equally with all the bacteria in the system; this makes them more vulnerable to the cheating by the non-producers. However, we will focus not only in the case of non-producers but also in the exploitation of the communication channel.

The amount of PGs released to the medium depends on the detected concentration of signal [7, 8], so bacteria will not produce PGs for low amounts of signal. In addition to that, production of PGs will saturate for high concentrations of signal. The signal molecule is an autoinducer (AI), meaning that it also induces its own production [3,7,8]. Then, a bacterium that detects a high concentration of signal also releases a high concentration of signal. This leads to the existence of two states, one «inactivated» with low production of AIs and no production of PGs, and an «activated» one with a high production of AIs and production of

PGs, stimulated by the signal [8]. We remark that the cost of producing AIs increases, due to a higher rate of production of the signal [7,9], but the cost of producing the PGs is still higher [7]. Note that even when the number of bacteria in the medium decreases, the production of public goods or the signal does not reduce immediately, since the molecules persist in the medium for a time that depends on the rate of degradation of the molecules [2,9]. This can be seen as a kind of hysteresis, because the number of individuals needed to go into the active state is higher than the number required to come back to the inactive [8].

In this point it is natural to wonder about the reason why bacteria regulate the PGs using QS. In biology, the adaptation to the medium is measured by the fitness of the species. Usually in biological systems there is no precise way of defining the fitness. However, bacteria are simple organisms and the fitness can be readily identified with the reproduction rate of the individuals [10]. Exoproducts have a metabolic cost for bacteria, meaning that they have to invest energy to produce the molecules [2,3,7], so an individual that has a high metabolic cost will reproduce slower. As we have said, public goods are costly to produce, and they only have a benefit for the population when the number of individuals is large enough, so if they were produced when the density of bacteria is low, the fitness of the system would be lower [2,7,11]. Theoretical studies [7] and experiments [11] show that having a quorum sensing regulation system allows bacteria to have a higher reproduction rate for low number of bacteria, avoiding the costly exoproducts when they do not give a benefit at the population level; once the population size is big enough, quorum sensing enables the production of PGs that increases the fitness of bacteria [11]. Even if most bacteria have a QS mechanism to regulate the production of costly molecules, some of them lack this system. Moreover, depending of the specific regulation and species, QS has to be accordingly tuned; as a result, the activation threshold at which the quorum sensing triggers a behaviour spans four orders of magnitude [11]. However, this demonstrates that QS is robust in the sense that it is an efficient control mechanism for a wide variety of situations [11].

As we have discussed above, a bacterium that starts producing public good has a high metabolic cost, but provides a benefit for the neighbours. Active bacteria enhance their signal production, inducing neighbouring cells to start their production of PGs, which will translate later into a benefit for the signal producer. In this context, the QS mechanism promotes a cooperative behaviour. However, this does not mean that QS is always used to promote cooperation between individuals [2]. One example is the cross talk, a particular type of QS used to share information between species of bacteria [2]. In this case, bacteria from species A may try to force some response from species B, so A gains a reproductive advantage over B [2].

1.1.2 Model organism

As we discussed above in this work we focus in bacteria whose public goods are regulated by a QS mechanism. In particular, our reference is the bacterium *Pseudomonas aeruginosa*, which can cause severe infections in humans [3,8]. It is responsible for relatively high percentage of infections, and it has direct influence in 38% of the deaths by pneumonia, as well as a high percentage of deaths in chronic patients of cystic fibrosis [8]. Some virulence factors of *Pseudomonas aeruginosa*, as elastase, are regulated by a quorum sensing mechanism [3,8]. Moreover, the production of biofilms, that grants bacteria protection over antibiotics, is also controlled by quorum sensing in these bacteria [2,8]. There is currently an intense research to find quorum sensing inhibitors that reduce bacterial virulence or biofilm production in the host, as strategies against bacterial antibiotic resistance [12]. However, quorum sensing is a mechanism difficult to inhibit [13].

Pseudomonas aeruginosa has two different pathways for quorum sensing, called *las* and *rhl*. Each one use a different kind of the N-acyl homoserine lactone molecule, called *LasI* and *RhlI*, for the *las* and *rhl* pathways, respectively [3,8]. We focus on the *las* pathway, which regulates, among other things, the production of LasB protease, also known as elastase. In a host, proteases are used to «digest» several proteins, that will be used as nutrients, enhancing the growth [3]. We will suppose that bacteria are in a medium where the elastase is required for the growth of the population, as the one used in the experiments by [3].

1.1.3 Exploitation by cheaters.

As we have discussed, PG production may be exploited by cheaters. In the case of PG production regulated by quorum Sensing, there is also a possibility for exploiting the communication channel. Diggle *et al.* [3] report two kinds of mutant cheaters for the *las* pathway are reported for *Pseudomonas aeruginosa*. The first one is known as *lasI* or *signal negative* cheater. This strain does not produce the autoinducer molecule, but is able to react to the signal producing public good [3]. Even when it produces PGs, providing a benefit to the neighbours, signal negative bacteria do not have the cost of producing the signal, so they outcompete the wild type strain. A population formed only by signal negative individuals will not produce PGs, since there is no signal to activate them. This strain will take advantage of the signal produced by the wild type, without producing it by themselves. On the other hand we have the *lasR* or *signal blind* mutant, which does not react to the signal in any way [3]. It simply produces a fixed amount of AIs, inducing the wild type to activate and start producing the public good. Note that this is an example where a strain can use the QS mechanism for non-cooperative purposes. This is in fact a direct way to exploit the communication channel, inducing the wild type strain to produce instead of doing it by themselves. As in the signal negative strain case, a population composed only by signal blind bacteria cannot generate public good.

Then, the fitness of the single population culture is higher for the wild type, which can produce both the AIs and the PGs, that provide a benefit to the population. The other two strains cannot generate the PGs, so they will not be able to grow as much as the wild type. However, if we are in a medium where the PG production makes no difference, both mutants will have higher fitness than the wild type. In addition to that, in a mixed population of wild type and mutants, the later will exploit the communication channel in order to get the benefit of the PG without paying the full costs [3]. Cheaters may outcompete the wild type, growing faster and leading wild type strain to extinction. However, in some cases the wild type is able to resist the invasion, leading to a coexistence between the two strains.

The exploitation of the communication channel and PG production by the cheaters leads to a well-known problem in cooperation called *tragedy of the commons* [3–5]. The fitness of the cheater depends on the initial number of cheaters in the population, so it is frequency dependent [3]. If the frequency of cheaters is very low, their fitness will be very high because they can get the benefit from the PGs without all the costs. However, if there if the frequency of cheaters is high, the concentration of PG in the medium will decrease, so the fitness of the whole population is reduced [2, 3]. If we compute the relative fitness between the cheaters and the wild type strains, we find that the relative fitness of the cheaters decreases with the increase frequency of cheaters, up to the point where cheaters grow slower than the wild type strain. This is called *negative frequency dependence.*

In this work we present an individual based stochastic model for a quorum sensing mechanism that regulates the production of public goods in *Pseudomonas aeruginosa.* The objective of the model is to understand how this regulation mechanism affects the growth of the strains in single and mixed populations. The model take into account both the AIs and PGs in an implicit way, via the local interaction with neighbouring bacteria. In the next Section we present the basic mathematical background that will be used later for the analysis of our model. In Chapter 2 we describe in detail the model, and analyse the mean field behaviour. Numerical simulations and results from the individual based model can be found in Chapter 3, showing results both from single and mixed populations, in the high and low diffusion limits to emphasize the role of the space. Finally, we present the conclusions in Chapter 4.

1.2 Mathematical tools

1.2.1 From individual reactions to master equations.

One of the mathematical basis of our model are master equations. These are differential equations that describe the evolution of the probability of finding the system in certain state, for systems that evolve «jumping» from one state to another [14–16]. Master equations provide an exact description of a stochastic process and allow us to derive the average macroscopic behaviour of the system starting from the microscopic dy-

namics. In this first section, we introduce some of the most important features of master equations and apply them to a case that will be of interest in order to understand our model: the logistic growth of a population.

The set of stochastic reactions we are going to use to illustrate this section is the birth and death by pairs. This model will be the basis of our model, and it is described by the reactions,

$$A \xrightarrow{k} 2A, \tag{1.1}$$

$$A + A \xrightarrow{\delta} A. \tag{1.2}$$

According to the reaction (1.1), a particle A can divide into two sibling particles, showing a simple reproduction behaviour, suitable for bacteria. On the other hand, (1.2) represents the interaction of two particles, that can lead to the death of one of them. The meaning of (1.2) is that two individuals may compete for a resource that is not available for everybody, resulting in the death of one of the individuals. k is the rate of reproduction, and δ the death rate at which one particle die because of the competition. Rates are measured in units of inverse of time, since they are probabilities per unit of time.

Then, imagine we have n particles in the system, at certain time t. We neglect the spatial component, and assume that the system is well-mixed in the volume V. Reactions (1.1) and (1.2) are for only one individual, so if we only had one particle, we can expect a reproduction event after k^{-1} units of time, in average. If we have two particles, the event would happen after $k^{-1}/2$ time in average. In general, for n particles we wait k^{-1}/n units of time for the next event, meaning that the reproduction event is happening with a total rate

$$\Omega\left(n \to n+1\right) = kn. \tag{1.3}$$

The interpretation of the second reaction is more complicated, because it includes the encounter of two particles. Individual events happen at rate δ. If there are n particles, the rate at which one of them is ready to trigger the reaction is, as before, $n\delta$. However, since we need another particle to interact with, the rate has to be proportional to the density of particles $(n-1)/V$ that remains in the system [14]. The term of the volume also indicates that particles that are too far away are not going to interact,

$$\Omega\left(n \to n-1\right) = \delta n \frac{(n-1)}{V} \equiv \delta n\left(n-1\right) \tag{1.4}$$

We are usually interested in the limit when $n \gg 1$ so $n-1 \simeq n$. Also, notice that we have redefined $\delta \equiv \delta/V$ as the effective death rate. Now we can do a brief analysis of the system. If n is very low, then the reproduction rate is higher and the system grows. However, when n grows, there is a point where both global rates become the same, so births and deaths balance each other and the system reaches a stationary state. This happens at the point

$$\Omega\left(n \to n+1\right) = \Omega\left(n \to n-1\right) \to N = \frac{k}{\delta}. \tag{1.5}$$

This quantity is the carrying capacity of the system, and represents the stable population size.

Next we write the master equation of this process, that will provide more details about the statistical properties of the system. The master equation describes how the probability $p\left(n,t\right)$ of having n particles at time t changes with time. To obtain it we assume that the system is at time t, and wait a very short time dt. If after this process, at time the system has n particles, three things could happen during the interval dt:

1. We had $n-1$ individuals and one reproduced. The probability is given by the total birth rate at t multiplied by the time interval, that is $k\left(n-1\right)dt$.

2. We had $n+1$ individuals, and then a death occurred. The probability of this event to happen is $\delta\left(n+1\right)^2 dt$.

3. We had n individuals and nothing happened: the probability for this event is the complementary probability of (1) and (2), so $(1-kndt)\left(1-\delta n^2 dt\right) = 1-kndt-\delta n^2 dt+\mathcal{O}\left(dt^2\right)$.

Notice that we use $n+1$, $n-1$ or n depending on how many particles we had in the time before the current state. Also, we assume that dt is short enough so we are not going to see two events happening in this interval of time, for example, two reproductions. We have neglected terms of order $\mathcal{O}\left(dt^2\right)$ due to this assumption. The probability of having n particles at time $t+dt$ is given by,

$$p\left(n,t+dt\right) = p\left(n-1,t\right)k\left(n-1\right)dt + p\left(n+1,t\right)\delta\left(n+1\right)^2 dt + p\left(n,t\right)\left[1-kndt-\delta n^2 dt\right]. \tag{1.6}$$

Equation (1.6) can be rewritten to obtain the temporal evolution of $p\left(n,t\right)$ in the limit $dt\to 0$, which is the master equation of this process,

$$\frac{p\left(n,t+dt\right)-p\left(n,t\right)}{dt} = \frac{\partial p\left(n,t\right)}{\partial t} = p\left(n-1,t\right)k\left(n-1\right) + p\left(n+1,t\right)\delta\left(n+1\right)^2 + p\left(n,t\right)\left[-kn-\delta n^2\right]. \tag{1.7}$$

Equation (1.7) allows us to compute of the evolution of the probability to have n particles at a time t. However, (1.7) relates $p\left(n,t\right)$ with $p\left(n+1,t\right)$ and $p\left(n-1,t\right)$, making the resolution of the equation complicated to do analytically. Some approaches can be followed in order to find solutions, such as the generating function method [14–16], but in any case the analytical solution to the problem is difficult to find.

1.2.2 Mean field and fluctuation analysis

Since it is often very difficult to find a complete analytical solution for the master equation, it is usual to compute the moments of the distribution instead. In the ideal case, we should obtain $\left\langle n^k\right\rangle$ for all values of k so the distribution is completely characterized. However, this approach is as difficult as solving the master equation. For this reason, we usually perform approximations that provide the lower order moments, which capture some of the most important properties of the distribution. In this section we compute the first two moments of the distribution associated to (1.13), in order to do mean field and fluctuations analysis.

First of all, we can write the master equation (1.7) in a more compact notation, if we include the ladder operator $E^{\ell}\left[f\left(n\right)\right] = f\left(n+\ell\right)$ and identify the global rates,

$$\frac{\partial p\left(n,t\right)}{\partial t} = p\left(n-1,t\right)\Omega\left(n-1\to n\right) + p\left(n+1,t\right)\Omega\left(n+1\to n\right) - p\left(n,t\right)\left[\Omega\left(n\to n+1\right)+\Omega\left(n\to n-1\right)\right] = \left(E^{-1}-1\right)\left[\Omega\left(n\to n+1\right)p\left(n,t\right)\right] + \left(E-1\right)\left[\Omega\left(n\to n-1\right)p\left(n,t\right)\right]. \tag{1.8}$$

In fact, it can be proved that any master equation characterized by a set of global rates $\{\Omega\left(n\to n-\ell\right)\}$ can be written, in a general way, as [14],

$$\frac{\partial p\left(n,t\right)}{\partial t} = \sum_{\ell}\left(E^{\ell}-1\right)\left[\Omega\left(n\to n-\ell\right)p\left(n,t\right)\right]. \tag{1.9}$$

Note that we only have the terms that corresponds to $\ell=+1,-1$ in the logistic equation. The form (1.9) of the master equation is useful because it allows the computation of moments of the distribution of n easily. The equations for the first two moments [14] are given by,

$$\frac{d\left\langle n\right\rangle}{dt} = -\sum_{\ell}\left\langle \ell\Omega\left(n\to n-\ell\right)\right\rangle, \tag{1.10}$$

$$\frac{d\left\langle n^2\right\rangle}{dt} = \sum_{\ell}\left\langle \ell\left(\ell-2n\right)\Omega\left(n\to n-\ell\right)\right\rangle. \tag{1.11}$$

In the case of the logistic equation, when we follow this procedure, we reach the following system of equations,

$$\begin{cases} \frac{d\langle n\rangle}{dt} = k\langle n\rangle - \delta\langle n^2\rangle, \\ \frac{d\langle n^2\rangle}{dt} = k\langle n\rangle + (2k+\delta)\langle n^2\rangle - 2\delta\langle n^3\rangle \end{cases} \tag{1.12}$$

System (1.12) usually needs an approximation to be solved, since it is not closed. Notice that, if we were interested only in the first moment, we would need equation for the second moment. The equation for second moment needs an expression for the third one, and so on. Since we do not want to solve the equations for all the moments, we have to cut at some point.

Let's say we want to keep only the first moment. Then, the usual procedure is to use the *mean field approximation*, that makes $\langle n^2\rangle = \langle n\rangle^2$. This is equivalent to assume that statistical fluctuations are negligible, since we are setting $\sigma^2(n) = \langle n^2\rangle - \langle n\rangle^2 = 0$. When we do this approximation in the first equation of (1.12), we obtain

$$\frac{d\langle n\rangle}{dt} = k\langle n\rangle\left(1 - \frac{\delta}{k}\langle n\rangle\right) = k\langle n\rangle\left(1 - \frac{\langle n\rangle}{N}\right), \tag{1.13}$$

where we have used the definition of the carrying capacity (1.5). Equation (1.13) is known in the literature as the logistic equation, and it is a widely known equation in population dynamics. The logistic equation is used to model a population that has an initial exponential growth, but saturates when $\langle n\rangle = N$. This can be seen by the computation of the fixed points of (1.13), that are

$$\langle n\rangle_1 = 0 \qquad \text{(unstable)}, \tag{1.14}$$

$$\langle n\rangle_2 = N \qquad \text{(stable)}. \tag{1.15}$$

This means that the population is going to grow until we reach the carrying capacity. Note that this kind of behaviour is compatible with the discussion we did before using the global rates, so we have been able to go from the microscopic point of view governed by the stochastic reactions to the description of a macroscopic variable such as the mean number of individuals. Although the logistic equation is non-linear, it has an analytical solution given by

$$\langle n(t)\rangle = \frac{n_0 N e^{kt}}{N + n_0\left(1 - e^{kt}\right)}, \tag{1.16}$$

where n_0 is the initial number of individuals. The fact that the logistic equation can be treated analytically will be very useful for the model parametrization later.

However, as we said, we are neglecting the fluctuations in this analysis. If we want to take into account the fluctuations, we have several options, such as the study of the Fokker-Planck or Langevin equations that are associated with the master equation [15, 16]. These equations can be derived using a large-size van Kampen expansion from the master equation, that allow us to do an analysis of the fluctuations [15, 16]. In contrast, in this case we study the fluctuations dropping the mean field approximation and including the second moment equation, i.e. solving the system (1.12). We still need an approximation for the third moment. One of the most usual is the Gaussian approximation, that assumes that we can use a relation valid for Gaussian variables in order to write the third moment in terms of the first two,

$$\left\langle n^3\right\rangle = 3\langle n\rangle\left\langle n^2\right\rangle - 2\langle n\rangle^3, \tag{1.17}$$

so the system of equations (1.12) is now closed, and has the form,

$$\begin{cases} \frac{d\langle n\rangle}{dt} = k\langle n\rangle - \delta\langle n^2\rangle, \\ \frac{d\langle n^2\rangle}{dt} = k_0\langle n\rangle + (2k+\delta)\langle n^2\rangle - 6\delta\langle n\rangle\langle n^2\rangle + 4\delta\langle n\rangle^3. \end{cases} \tag{1.18}$$

It is interesting to analyse also the fixed points of the system, in order to compare with the mean field approximation. As before, we have $\langle n\rangle = \langle n^2\rangle = 0$ which is an unstable fixed point. In addition to that,

two new fixed points appear,

$$\begin{cases} \langle n \rangle_{\pm} = \frac{3}{4}\frac{k}{\delta} \pm \frac{1}{4}\sqrt{\frac{k}{\delta^2}(k-8\delta)} = N\left(\frac{3}{4} \pm \frac{1}{4}\sqrt{1-\frac{8}{N}}\right) \\ \langle n^2 \rangle_{\pm} = \frac{3}{4}\frac{k^2}{\delta^2} \pm \frac{1}{4}\frac{k}{\delta^2}\sqrt{\frac{k}{\delta^2}(k-8\delta)} = N^2\left(\frac{3}{4} \pm \frac{1}{4}\sqrt{1-\frac{8}{N}}\right) \end{cases} \tag{1.19}$$

The point $n_+ \equiv \left(\langle n \rangle_+, \langle n^2 \rangle_+\right)$ is a stable fixed point, which corresponds to the new carrying capacity $\overline{N}$, while n_- is a saddle point. We can compute the carrying capacity $\overline{N}$ in terms of the old carrying capacity N, using the fact that $N \gg 1$ so we can Taylor expand,

$$\overline{N} = N\left(\frac{3}{4} \pm \frac{1}{4}\sqrt{1-\frac{8}{N}}\right) \simeq N - 1 + \mathcal{O}\left(N^{-1}\right), \tag{1.20}$$

$$\left\langle n^2 \right\rangle_+ \simeq N^2 - N + \mathcal{O}(1) = N(N-1) + \mathcal{O}(1). \tag{1.21}$$

We see that the correction to the carrying capacity due the introduction of the fluctuations is very low. Since we have now an expression for the second moment at the fixed point, it is also possible to evaluate the variance

$$\sigma^2(n_+) = \left\langle n^2 \right\rangle_+ - \overline{N}^2 \simeq N. \tag{1.22}$$

Finally,. using equation (1.22), we can evaluate the dependence of the fluctuations on the carrying capacity of the system,

$$\varepsilon \sim \sqrt{\frac{\sigma^2(n)}{N}} \simeq \frac{1}{\sqrt{N}}. \tag{1.23}$$

Relation (1.23) is typical in statistical mechanics, and tell us that the fluctuations go to zero with the inverse square root of the number of particles. In this case, the fluctuations decay with the number of particles that correspond to the stationary state of the system. Equation (1.23) is very useful since it tell us that if we increase enough the carrying capacity, fluctuations will be very low. Since we will have $\varepsilon \simeq 0$, the mean field approximation holds, and we will be able to take advantage of all the analytical results of the logistic equation, such as its solution (1.16).

1.2.3 The Gillespie algorithm

The Gillespie algorithm, also called residence time algorithm, is an algorithm that allows the exact computation of stochastic reactions numerically [14,17,18]. First we explain how to use this method for a Markovian process, using the particle point of view [14], following [17, 18]. We are going to compute the probability density $f(\mu, \tau|n, t)$ that represents the probability that, given n individuals at time t, the process μ will happen in the time interval $[t, t+\tau]$. In general, the total number of reactions is $M = n \cdot n_r$, where n_r is the number of reactions per individual. The index μ runs for every possible reaction $\mu = 1, 2, \ldots, M$.

Let's find first an explicit expression for $f(\mu, \tau|n, t)$. To do that, we start computing the probability of having the reaction μ in the time interval $[t, t+\tau]$. We divide the interval into m subintervals of length $\Delta\tau = \tau/m$. Then, $f(\mu, \tau)\, d\tau$ is the probability that none of the reactions happened in any of the subintervals except the last one, so it is

$$f(\mu, \tau|n, t)\, d\tau = \left(1 - \Delta\tau \sum_{j=1}^{M} k_j\right)^m \cdot \left(k_\mu \Delta\tau\right). \tag{1.24}$$

where k_μ is the global rate of reaction μ. In general the term $\sum_j k_j = a$ which accounts for the contribution of all the reactions of the system is called the total rate [17]. Now, taking the limit $\Delta\tau \to 0$ (i.e. making the

number of subdivisions $m \to +\infty$) we find that,

$$f(\mu, \tau | n, t) = k_\mu e^{-a\tau} \equiv \left(a e^{-a\tau}\right) \left(\frac{k_\mu}{a}\right). \tag{1.25}$$

We see that $f(\mu, \tau | n, t)$ is the product of other two probability density functions: the probability density function of having the next event after a time τ, which nothing but an exponential, and the probability density of having the process μ at that time, which assign a constant probability proportional to k_μ each reaction. To draw random numbers following this probability distributions, we generate two uniform random numbers ξ_1 and ξ_2 between 0 and 1 and then compute τ and μ following their distributions,

$$\tau = -\frac{\log(\xi_1)}{a}, \tag{1.26}$$

$$\sum_j^\mu k_j \leq \xi_2 < \sum_j^{\mu+1} k_j. \tag{1.27}$$

These will say us which reaction we choose and at which time happens since last event. For example, in the case of the logistic equation, we have to take into account that $a = \Omega(n \to n-1) + \Omega(n \to n+1) = kn + \delta n^2$. The rate a has to be computed every iteration, since the total rates depend on the number of particles n. In summary, to use the Gillespie algorithm, we follow the steps [17,18]:

1. Get uniform random number between 0 and 1.
2. Compute τ and μ.
3. Execute reaction μ, adding or deleting particles from the system.
4. $t = t + \mu$. Start again.

1.3 Spatial degrees of freedom

1.3.1 Bacterial motility

In our model, bacteria will move following a Brownian motion, with constant diffusion coefficient D. As we know, the position of a particle that moves with pure diffusion is distributed as a Gaussian function [14,15], where

$$\langle \vec{x}(t) \rangle = 0,$$
$$\left\langle |\vec{x}(t)|^2 \right\rangle \equiv \left\langle \ell^2 \right\rangle = 2Dt,$$

so the fluctuations increase linearly with t. We can perform numerically a diffusive random walk drawing uniform random numbers for the displacement, $\vec{x}(t_{n+1}) = \vec{x}(t_n) + \vec{\xi}_n$. Since the time step is not fixed, we have to be careful because the individual will move with different velocity each step, leading to non-constant D [19]. We must make the jump length variable in order to fix the diffusion coefficient. The probability of having a change after a time τ is given by (1.25), which is an exponential distribution. The mean $\overline{\tau}$ of this probability distribution is given by

$$\overline{\tau} = \frac{1}{a}, \tag{1.28}$$

so, when bacteria reproduce, the total rate a increases and the time step decreases, i.e. there is less time between events.

To address the problem of non-constant diffusion produced by the variable timestep, we will also generate a pair or random numbers, one for the jump length ℓ and another for the direction angle [19]. The angle can be sampled using a uniform distribution in $[0, 2\pi]$, while the jump length will be sampled using a Gaussian distribution

$$p(\ell) = \frac{2}{\tilde{\ell}\sqrt{2\pi}} \exp\left(-\frac{\ell^2}{2\tilde{\ell}^2}\right), \tag{1.29}$$

The variance of the jump length is given by $\langle \ell^2 \rangle = \tilde{\ell}^2$. In a Brownian movement, this value is related with the diffusion coefficient, so we can estimate the value of D making $t = \overline{\tau}$. Finally, we need to scale eq. (1.29) accordingly. To do this, we choose to fix the value of D so bacteria always diffuses the same in every time step [19]. This is done by computing

$$\tilde{\ell} = \sqrt{\frac{2D}{a}}, \tag{1.30}$$

and we use this value in equation (1.29). Using this method we ensure we are taking a constant diffusion coefficient D and that bacteria moves via pure diffusion correctly.

1.3.2 Reaction-Diffusion equations

We now want to consider the role of the space into our mean field equation. We can take the mean field equation (1.13) and make $n = n(\vec{x})$. In each realization of the individual based, discrete model, the field that gives the number of particles is given by

$$n(\vec{x}) = \sum_{j=0}^{n} \delta(\vec{x} - \vec{x}_j) \tag{1.31}$$

where $\vec{x}_j$ is the position of every particle. However, when we average over the collectivity of realizations, the value of $n(\vec{x})$ becomes a real number for each position, whose value will be proportional to the probability to find a particle in $\vec{x}$. This is a natural way to define the density field,

$$\rho(\vec{x}, t) = \frac{\langle n(\vec{x}, t) \rangle}{V}. \tag{1.32}$$

When we do this substitution in (1.13), we get

$$\frac{\partial \rho(\vec{x}, t)}{\partial t} = D\nabla^2 \rho(\vec{x}, t) + k\rho(\vec{x}, t)\left(1 - \frac{\rho(\vec{x}, t)}{\rho_N}\right) \tag{1.33}$$

where we have defined $\rho_N = N/V$ as the density corresponding to the carrying capacity. In addition to that, note that we have added the diffusion term, whose contribution comes from the Brownian movement of the particles [1]. These equations are known as reaction-diffusion problems, and they appear in diverse applications such as chemical reactions, or transport in fluid dynamics [20, 21].

The behaviour of reaction-diffusion systems is very rich, since they are able to exhibit oscillations, excitability, chaos and pattern formation. The most popular example for chemical reactions is the Belóusov-Zhabotinsky reaction. In general, a reaction-diffusion system with N different chemicals or species is characterized by a system of equations of the form

$$\partial_t \rho_j(\vec{x}, t) = D_j \nabla^2 \rho_j(\vec{x}, t) + k_j f(\vec{\rho}). \tag{1.34}$$

It turns out that equation (1.33), is one of the most illustrative models that can be studied in the problem of reaction-diffusion [20]. In order to understand (1.33), we are going to use the stochastic reaction

$$A + R \to 2A, \tag{1.35}$$

[1] If we start the derivation from microscopic description, as in Section 1.2.2, this term comes from the fact that the number of particles $n(\vec{x}, t + dt)$ come from two different contributions now: the birth or death of individuals as before, and the flux of particles that enter or leave an area defined by $\vec{x} + d\vec{x}$. This gives a term that leads to the spatial derivative in the limit $d\vec{x} \to 0$.

that at the mean field level is described also by (1.33). However, fluctuations are different, as described later. Reaction (1.35) represents the interaction of two chemical species, A and R. If we introduce a low concentration of A in a volume that contains R, it starts to react producing more A until the concentration of R reaches zero, so we cannot generate more A in that point, leading to a saturation. Then, A is going to grow up to the maximum possible concentration ρ_N, eliminating all the R available in the point $\vec{x}$. However, at the same time, A will diffuse and start the reaction in the nearby points. At the end, what we have is the propagation of a front. In general, it can be demonstrated, given some conditions for $f(\vec{\rho})$, that we have front propagation [20] with speed given by

$$v_f = 2\sqrt{kDf'(0)}. \tag{1.36}$$

For equation (1.33), the front velocity reads

$$v_f = 2\sqrt{kD}. \tag{1.37}$$

This front will be perfectly circular in the mean field equation, but not in the stochastic realizations, in which we will have a rough surface. Statistical properties of these surfaces can also be studied, but they are not of interest in the context of this work. Equation (1.33) have been solved numerically in order to see the front propagation, as depicted in Figure 1.1.

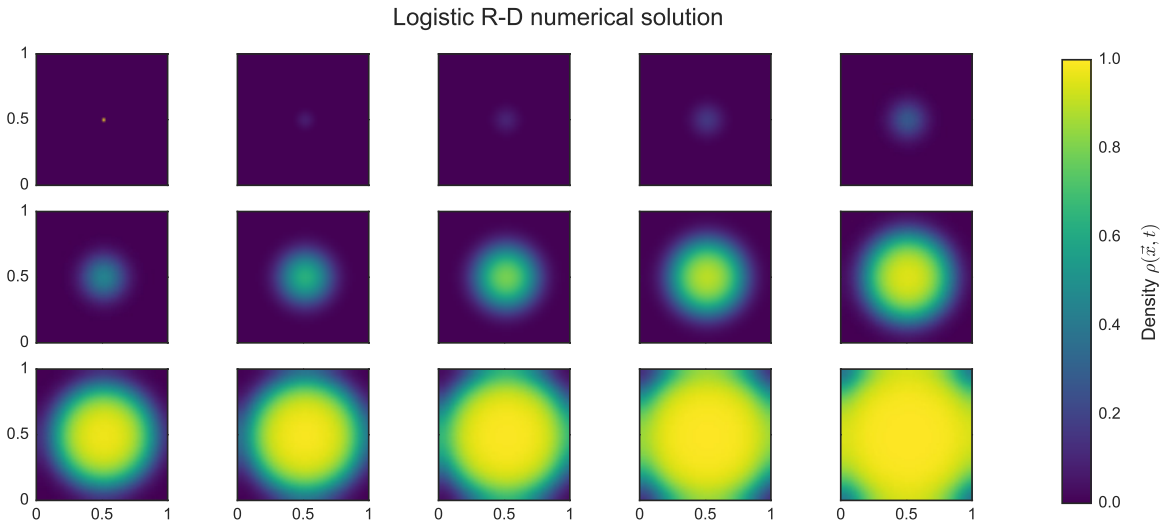

Figure 1.1: Evolution from $t_0 = 0$ to $t_f = 15$ (in arbitrary units) of equation (1.33), for initial condition $\rho(\vec{x},0) = \delta(x - L/2)\,\delta(y - L/2)$ and periodic boundary conditions. Parameters are $k = 1.0$, $D = 10^{-3}$ and $\rho_N = 1.0$. We can see that the solution is a front that propagates. The front reaches the wall at $t \sim 8$, which is compatible with the value got by (1.37).

In the context of population dynamics, the interpretation of (1.35) means that species A can grow if there are resources available in the system, represented by R. We stress that the mean field equation of (1.35) is (1.13), so in this sense it is equivalent to our birth and death reactions (1.1) and (1.2). The main difference comes in the fluctuations. Suppose we are near the carrying capacity, so $n = N$. For (1.1) and (1.2), we have the same probability to have a birth of a death, so the population can either grow or decrease. This deviation will not last very long, since if n starts to be higher than N then the rate of death increases, leading the system again to $n = N$. However, in (1.35), the state $n = N$ is an absorbing state, since once the system exhausts

reactant R, it reaches the equilibrium and no more changes happen. In addition to that, the concentration of A never decreases, in contrast with (1.2). We have done a comparison of the differences of our complete model for bacteria using also this kind of reaction in Appendix A.

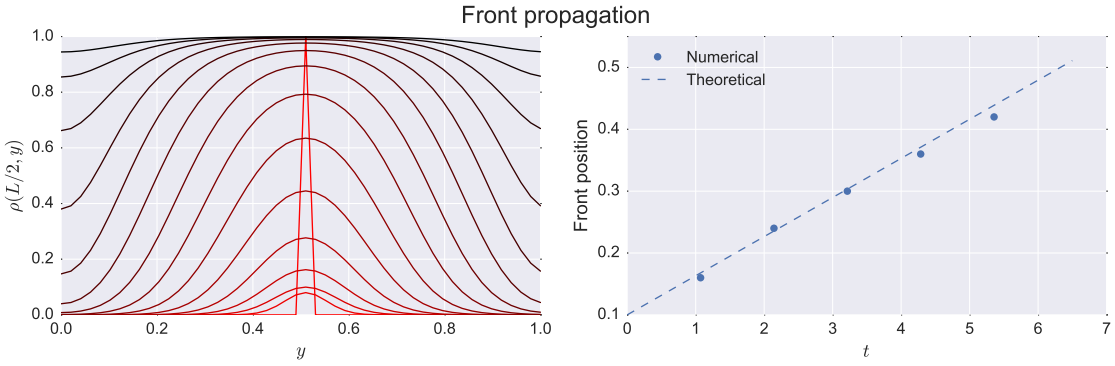

Figure 1.2: Left: concentration $\rho\left(\frac{L}{2}, y\right)$ for the different snapshots shown in Figure 1.1. Red correspond to $t = t_0$ and black to $t = t_f$. Right: comparison of the obtained position of the front from numerical simulations and (1.37). Despite numerical errors, we see that the tendency is very similar.

2. Analysis of a model of Quorum Sensing.

2.1 Description of the model.

We consider an individual based stochastic model for a public good production controlled by quorum sensing in a bacterial population. In this Section we develop a model given all the biological considerations discussed in Section 1.1. The model includes three important processes:

1. Stochastic reactions for birth and death processes, that will account for cell reproduction.
2. Interaction between the individuals due to the production of autoinducers and public good molecules.
3. Space, in which bacteria will move and interact.

Let us start explaining in detail the system. We will focus now in the wild type bacteria, because produces both the AI and PG molecules. We have a system of n bacteria, where each individual j has its own growth rate k_j. This growth rate will depend on the metabolic costs associated to the production of autoinducers and public good, as well as the benefit due to presence of public good. Then, the stochastic reactions for the system have the form

$$A_j \xrightarrow{k_j} A_j + A_{n+1}, \tag{2.1}$$

$$A_j + A_i \xrightarrow{\delta} A_j, \tag{2.2}$$

where the subindices are used to stress the fact that each bacterium has each own growth rate. The interaction of a bacterium with other individuals is taken in account as an effective way in k_j. An individual which detects a high concentration of public good will have a higher reproduction rate, doubling faster than their neighbours. We consider that the benefit of the PG increases the growth rate proportionally to the amount of detected public good by individual j, n_j^{pg}.

As we discussed in the introduction, bacteria produce an autoinducer molecule. The autoinducer production has a cost, q. When the individual detects enough concentration of autoinducer, it activates, and it starts producing more signal, as well as public goods. As a consequence, an active individual has a higher cost for the signal production, $\beta_{qs}q$, and a cost for the PG production $\beta_{pg}q$. The total cost is modeled as $\beta = \beta_{qs} + \beta_{pg}$. Taking this in account, the expression of the growth rate can be identified with the benefit provided by the PGs minus the costs of production of autoinducers and public goods [4,5],

$$k_j = k_0 - q + \alpha n_j^{pg} \qquad \text{(inactive)}, \tag{2.3}$$

$$k_j = k_0 - \beta q + \alpha n_j^{pg} \qquad \text{(active)}, \tag{2.4}$$

where inactive or active are the two possible states for the bacteria, triggered by the quorum sensing mechanism, and α is a proportionality constant that weights the fitness benefit of PG. The next step is to explain how a bacteria changes between inactive or active states.

In order to do that, we model the interactions due to the QS mechanism. We are going to use two radii, r_{qs} and r_{pg}, for the detection of the AIs and the PGs, respectively. This is because we want to take into account the diffusion of the molecules implicitly, so we are not going to track each single molecule. We compute the number of bacteria at a distance less than r_{qs} (for AI) or r_{pg} (for PG) from an individual. When we count bacteria inside the area defined by r_{qs}, we assign a different weight for active and inactive neighbours, since active bacteria produce a higher amount of signal. Let's say a bacterium j has detected n_j^{inac} inactive bacteria and n_j^{act} active bacteria. This is clarified in Figure 2.1. Then, the amount of autoinducer detected is given by

$$n_j^{qs} = n_j^{inac} + \beta_{qs} n_j^{act}, \tag{2.5}$$

As we have said, active bacteria produces β_{qs} times more AI so it is weighted by this amount. An individual will become active if $n_j^{qs} \geq u_{qs}$, where u_{qs} is the activation threshold. Once the bacterium is active, it will stay in this state at least during a certain time. For the timescales we are working with, we assume that the inactivation will not happen and once a bacteria is active it will remain in this state [8, 9]. Active bacteria have enhanced AIs production, as well as PGs production.

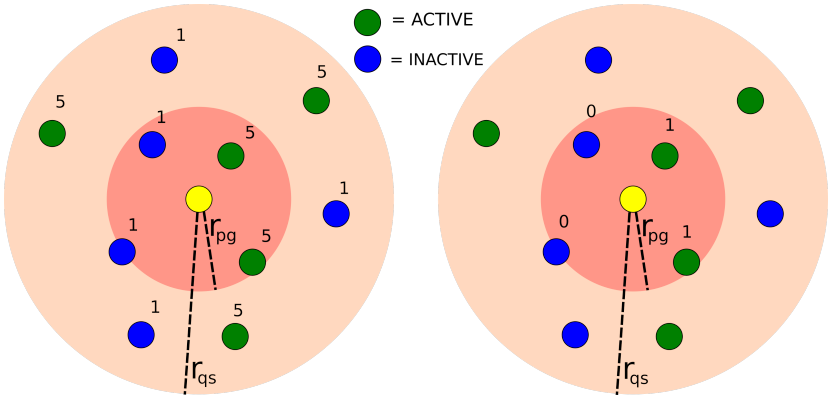

Figure 2.1: Example of computation of detected AIs and PGs. The molecules are implicit in the model, and the detection is based in neighbouring cells. In left, yellow individual detects AIs. Both inactive and active individuals inside the area defined by r_{qs} contribute, since they produce AIs, but they are weighted in a different way. In right, the yellow individual detects PGs. Only active bacteria produce PGs, so we count active individuals in the area defined by r_{pg}.

The public good is less diffusive than the signal, given its size[1]. For this reason, the radius of detection of PG is lower than the radius of detection of the signal molecule, $r_{pg} < r_{qs}$. It is interesting to stress that in this model we have two interaction scales, one for the communication channel and one for the PG. The effect of different interaction scales has been studied in the literature [22]. However, in our problem both scales have a similar range so we do not expect a great relevance in the dynamics of the system. In our model the amount of public good detected is the number of active individuals inside the area defined by r_{pg}. However, there is a limit in the benefit provided by the public goods, i.e., there is certain threshold u_{sat} such that if $n_j^{pg} \geq u_{sat}$ there is no difference in the benefit received. For this reason, the PGs detected can be written using the Heaviside unit step function[2],

$$n_j^{pg} = n_j^{act} \theta\left(u_{sat} - n_j^{act}\right) + u_{sat} \theta\left(n_j^{act} - u_{sat}\right), \tag{2.6}$$

[1]The signal is a small molecule, while the public good is a protein. The size of the PGs is much larger than the AIs, so we assume that the diffusivity of the signal is larger.

[2]defined as $\theta(x - a) = 1$ if $x \geq a$, and 0 otherwise.

which is equivalent to the piecewise system

$$n_j^{pg} = \begin{cases} n_j^{act} & u_{sat} \geq n_j^{act} \\ u_{sat} & u_{sat} < n_j^{act}. \end{cases} \tag{2.7}$$

Even when the piecewise system (2.7) looks clearer, notation like 2.6 will be useful. When the spatial dependence is included, it is not possible to write the Heaviside step functions as a piecewise system, since their arguments will depend on local concentrations.

At this point we have described the stochastic reactions, as well as the individual growth rates, at a microscopic level. The movement in space will be characterized by the method indicated in Section 1.3.1, in a two dimensional space with periodic boundary conditions. It is worth to notice that real movement of bacteria is more complex than simple diffusion. If the concentration of nutrients is high enough, bacteria simply diffuse. However, if there is a low amount of nutrients, they will try to follow the gradient of concentration of chemicals. In many species this movement is done via a flagella motor that can be mounted or eliminated [10]. This motor allows bacteria to travel large distances, so the movement is far from being a Brownian motion. There are more realistic models for bacteria movement that could be implemented. Although this may be an important contribution, we neglect it for the sake of simplicity.

2.2 Mean-field description.

2.2.1 Homogeneous space.

Once we have described the microscopic behaviour of the individual based model, we focus now in the macroscopic description. In this section we obtain expressions for the average number of particles $\langle n \rangle$ using the mean field approximation, as well as the system for first two moments using the Gaussian approximation.

The total rates of equations (2.1) and (2.2) are given by

$$\Omega\left(n \to n+1\right) = \sum_{j=1}^{n} k_j, \tag{2.8}$$

$$\Omega\left(n \to n-1\right) = \frac{\delta}{L^2} n\left(n-1\right) \simeq \delta n^2. \tag{2.9}$$

We now analyse in detail the term $\Omega\left(n \to n+1\right)$. In the case of the logistic equation (1.13), since the rates k_j are equal, this can be written as nk. If we use in expressions (2.3) and (2.4) of growth rates for inactive and active particles, we can write,

$$\sum_{j=1}^{n} k_j = nk_0 - \left(n^{inac} + \beta_{qs} n_j^{act}\right) q + \alpha \sum_{j=1}^{n} n_j^{pg}, \tag{2.10}$$

where n^{inac} and n^{act} are the total numbers of inactive and active bacteria in the population, respectively. The total number of particles is $n = n^{inac} + n^{act}$. We still have to simplify the expression for the detected amount of public good detected. Equation (2.6) says that the detected PG by individual j only depends on the number of active particles detected, n_j^{act}.

We are interested in the average values. $\left\langle n_j^{act} \right\rangle$ can be obtained multiplying the number of active particles in the system by the probability of having an individual inside the area defined by r_{qs},

$$\left\langle n_j^{act} \right\rangle = \left\langle n^{act} \right\rangle p\left(\vec{x}_j, \{\vec{x}_i\}\right) \equiv \left\langle n^{act} \right\rangle p\left(j\right), \tag{2.11}$$

where $p(j)$ is the probability to detect a neighbour, given that we are focusing in one bacterium located in $\vec{x}_j$, and the location of the other particles in the system. If spatial structure of the system is ignored, the probability to detect a neighbour is the same for each individual so it has a constant value,

$$p(j) \equiv p = \pi \left(\frac{r_{pg}}{L}\right)^2. \tag{2.12}$$

Moreover, the number of detected AI or PG molecules is exactly the same for every cell, since the system is homogeneous in space. As a consequence, all the active individuals detected $\left\langle n_j^{act} \right\rangle$ are equal. This implies that all individuals detect the same amount of PG,

$$\left\langle n_j^{pg} \right\rangle = \langle n^{pg} \rangle \quad \forall j \tag{2.13}$$

Finally, we only have to write the mean number of active particles in the system as,

$$\left\langle n_j^{act} \right\rangle = \langle n^{act} \rangle \, p \equiv p^{act} p \, \langle n \rangle \,, \tag{2.14}$$

where we have written the total number of active individuals as the probability of becoming active multiplied by the total number of particles. The probability of being active p^{act} in the general case depends on the number of particles in the system and in its spatial distribution. We will discuss more in deep the activation probability in Section 2.2.2. In order to simplify the notation we can group the product $p^{act} p = P$, which is the probability of having and neighbour *and* that it is activated.

The value of n^{pg} given is limited by u_{sat} as we discussed, in equation (2.6), which was for each particle j. However, since we do not have any spatial structure the number of detected active particles is the same for every cell in the population,

$$\langle n^{pg} \rangle = \left\langle n^{act} \right\rangle p\theta \left(u_{sat} - \left\langle n^{act} \right\rangle\right) + u_{sat}\theta \left(\left\langle n^{act} \right\rangle - u_{sat}\right). \tag{2.15}$$

Note that we have also taken averages inside the arguments of the step functions. This is because all the individuals detect the same number of active particles in average. Now we have to insert (2.15) and (2.14) in the average value of the total rate $\Omega(n \to n+1)$. For the moment we assume all bacteria in the system are active, making $n^{inac} = 0$. This will make the notation easier, and allow us to recover the case with inactive individuals at the end. The average of the global rate reads

$$\langle \Omega(n \to n-1) \rangle = \sum_{j=1}^{n} \langle k_j \rangle = \langle n \left(k_0 - \beta q + \alpha n^{pg}\right) \rangle \,. \tag{2.16}$$

where we can identify

$$k(n) = k_0 - \beta q + \alpha n^{pg}. \tag{2.17}$$

with the average rate of the population, which is also given by,

$$k(n) = \frac{1}{n} \sum_{j}^{n} k_j.$$

Note that this average rate is performed over the growth rates of the individuals of a stochastic realization, and not over the realizations. Finally, we take the average of the detected number of public good considering that the active number of bacteria can be written using $n^{act} = n p^{act}$. This gives us the expression of the first moment,

$$\frac{d\langle n \rangle}{dt} = \langle n \rangle \left[k_0 - \beta q + \alpha \left(\langle n \rangle \langle P \rangle \, \theta \left(u_{sat} - \left\langle n^{act} \right\rangle\right) + u_{sat}\theta \left(\left\langle n^{act} \right\rangle - u_{sat}\right)\right)\right] - \delta \langle n \rangle^2 \,, \tag{2.18}$$

where we have applied the mean field approximation $\langle n^2 \rangle = \langle n \rangle^2$. Note that this also implies that $\langle P(n)\, n \rangle = \langle P(n) \rangle \langle n \rangle$. Looking at the step functions, we see that this is in fact a piecewise defined system. Writing it as a piecewise system, and regrouping the terms, equation (2.18) becomes,

$$\frac{d\langle n \rangle}{dt} = \langle n \rangle (k_0 - \beta q) \left[1 - \frac{\delta - \alpha \langle P \rangle}{k_0 - \beta q} \langle n \rangle\right] \left\langle n^{act} \right\rangle \leq u_{sat}, \tag{2.19}$$

$$\frac{d\langle n \rangle}{dt} = \langle n \rangle (k_0 - \beta q + \alpha u_{sat}) \left[1 - \frac{\delta}{k_0 - \beta q + \alpha u_{sat}} \langle n \rangle\right] \left\langle n^{act} \right\rangle > u_{sat}. \tag{2.20}$$

Note that since $\left\langle n^{act} \right\rangle < u_{sat} \implies \langle n \rangle \left\langle p^{act}(n) \right\rangle < u_{sat}$, so the separation between the two parts of the system is done in terms of n. When the system is above the threshold is described by (2.20), which is a logistic equation with growth rate $k_{eff} = k_0 - \beta_{qs} q + \alpha u_{sat}$ and carrying capacity $N_{eff} = k_{eff}/\delta$. Equation (2.19), however, is more complicated, since the benefit provided by the PG is greater than the death rate, ($\alpha \langle P \rangle > \delta$) and $\langle n \rangle \to +\infty$ exponentially as t grows. The dynamics of a system which has all the individuals active is the following: they start growing following (2.19). They do not saturate, so the population will grow until $n^{act} = u_{sat}$. At this point individuals do not obtain more benefit from the production of public goods, and the dynamics is governed by (2.20), which will saturate at the carrying capacity of the system, N_{eff}.

We neglected inactive particles from our analysis. We can take them in account easily, since the only difference is that we have to include the differences in the cost between an active and an inactive bacterium. In a mean field approximation, where the population is well mixed, once one bacterium crosses the threshold, in average all bacteria have enough neighbours to cross it, so we can approximate the probability of activation $\left\langle p^{act}(n) \right\rangle \simeq \theta\left(u_{qs} - \langle n \rangle p\right)$. When we do it, we find an equation for the inactive bacteria,

$$\frac{d\langle n \rangle}{dt} = \langle n \rangle (k_0 - q) \left[1 - \frac{\delta}{k_0 - q} \langle n \rangle\right] \langle n \rangle \leq u_{qs}/p, \tag{2.21}$$

which is a logistic equation with different growth rate. The threshold u_{qs} will be reached before the carrying capacity. Using the approximation we did for p^{act}, below the threshold $n < u_{qs}$ all individuals are going to be inactive, while above the threshold all the individuals will be active. Then, the population grows until the number of bacteria is enough to reach the threshold u_{qs} and then the system becomes the one with all bacteria active.

Once we have done the complete derivation, it is easy to see what is the behaviour of the system:

1. In the first part all bacteria are inactive, and the population grows following a logistic growth with $k = k_0 - q$. Governed by (2.21)

2. In the second stage bacteria are active and the system has an increasing carrying capacity, so the population grows exponentially, governed by (2.19). In this case the difference with the standard logistic equation (1.13) is the substitution $-\delta \to -\delta + \alpha \langle P \rangle$, with $\alpha \langle P \rangle > \delta$.

3. The third part happens after PG detection saturation, and the system follows a logistic with growth rate $k = k_0 - \beta q + \alpha u$. It follows (2.20).

Since (2.19), (2.20) and (2.21) can be easily identified with the logistic equation, it is straightforward to write the equations for the second moment $\langle n^2 \rangle$, identifying the expressions for k and δ and substituting them. This allows us to do an analysis of the fluctuations in the model. The complete system of equations is given by,

$$\begin{cases} \frac{d\langle n \rangle}{dt} = (k_0 - q) \langle n \rangle - \delta \langle n^2 \rangle \\ \frac{d\langle n^2 \rangle}{dt} = (k_0 - q) \langle n \rangle - 2\delta \langle n^3 \rangle + \langle n^2 \rangle (2(k_0 - q) + \delta) \quad \langle n \rangle \leq u_{qs}/p \end{cases} \tag{2.22}$$

$$\begin{cases} \frac{d\langle n \rangle}{dt} = (k_0 - \beta q) \langle n \rangle + (\alpha p - \delta) \langle n^2 \rangle \\ \frac{d\langle n^2 \rangle}{dt} = (k_0 - \beta q) \langle n \rangle + 2(-\delta + \alpha p) \langle n^3 \rangle + \langle n^2 \rangle (2(k_0 - \beta q) + \alpha p + \delta) \quad u_{qs}/p \leq \langle n \rangle \leq u_{sat}/p^{act} \end{cases} \tag{2.23}$$

$$\begin{cases} \frac{d\langle n\rangle}{dt} = \left(k_{eff} + \alpha u\right)\langle n\rangle - \delta\langle n^2\rangle \\ \frac{d\langle n^2\rangle}{dt} = \left(k_0 + \alpha u\right)\langle n\rangle - 2\delta\langle n^3\rangle + \langle n^2\rangle\left(2\left(k_{eff} + \alpha u\right) + \delta\right) \quad \langle n\rangle \geq u_{sat}/p^{act} \end{cases} \tag{2.24}$$

The first interesting thing about this system is that it has slope discontinuities. Such discontinuities are very small and they can be seen in the solution for $\langle n(t)\rangle$. However, the change from (2.23) to (2.24) is very clear in the expression of the variance, since during the exponential growth phase variance increases, until it reaches the last phase, where it stabilizes. Then it goes to the value given by equation (1.22), for the effective carrying capacity of the system. A numerical solution of the mean-field equations can be seen in Figure 2.2.

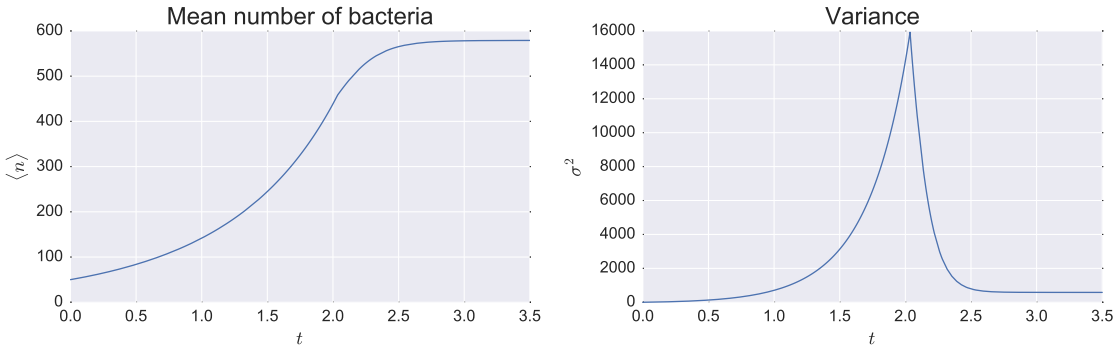

Figure 2.2: Numerical solution of a system composed only by active bacteria, as defined by (2.23) and (2.24). The discontinuity in the derivative between the two equations can be clearly seen in the variance. Parameters used (arbitrary units) were $k_0 - \beta q = 1.0$, $\delta = 10^{-3}$, $r_{qs} = r_{pg} = 1/6$, $L = 1$, $\alpha = 0.12$, $u_{qs} = 20$, $u_{sat} = 40$.

2.2.2 Spatial analysis.

In the previous section, we have discussed the mean field approximation, assuming that fluctuations are negligible due to $n \to +\infty$ and that the system is homogeneous. In this section we relax the condition of homogeneous space, so $\langle n(t)\rangle \equiv \langle n(\vec{x},t)\rangle$. However, we still neglect the fluctuations of the system. We will find an equation for the mean field density $\rho(\vec{x},t)$, as well as an expression for the probability of activation of a bacterium. The derivation of the mean field equation in this case is analogous as the before. The main difference now is that the relation $n_j^{pg} = n^{pg}$ for all j is no longer valid, because the spatial structure of the system is not homogeneous anymore.

We work now with the density of particles of the system, $\rho(\vec{x},t) = \langle n(\vec{x},t)\rangle / V$. The probability of detecting a bacteria in some area at time t is going to come from integrals of the probability density function of finding a particle, which is proportional to the density,

$$f(\vec{x},t) = \frac{\rho(\vec{x},t)}{\int_V d\vec{x}\rho(\vec{x},t)}. \tag{2.25}$$

Using (2.25), the probability of finding a neighbour inside a radius r , which was a constant before, is now a function of the position of each individual,

$$p(j) = \int_{\Omega_j} f(\vec{x},t)\, d\vec{x}, \tag{2.26}$$

where Ω_j is a circle or radius r centered in $\vec{x}_j$. The total number of particles in the system in one stochastic realization is going to be given now by $n(t)$. This can be obtained as the integral of the density. We remark that $\rho(\vec{x},t)$ is used for the mean field average, so its integration gives $\langle n(t)\rangle$.

In a next step, we have to find an expression for the activation probability. In order to do it, let's think that no particle in the system is active, and that the threshold for activation is u_{qs}. This means that an individual will need to detect u_{qs} or more particles in order to activate. Since the probability for the individual j to find a single individual is $p(j) \equiv p$, the probability of finding u_{qs} bacteria out of a system of n is given by the binomial distribution

$$p_b(n, u_{qs}) = \binom{n}{u_{qs}} p^{u_{qs}} (1-p)^{n-u_{qs}}. \tag{2.27}$$

Since we want to have u_{qs} or more particles, we have to sum all the binomial distributions starting with u_{qs} individuals, up to the point where the individual has detected the n particles,

$$p^{act}(j) = \sum_{i=u_{qs}}^{n} p_b(j, i). \tag{2.28}$$

However, this is correct only if every cell is inactive. Consider that we have only one active bacteria in the whole system. Then there is a new contribution to $p^{act}(j)$, because we can find at least u_{qs} inactive bacteria, *or* at least $u_{qs} - \beta_{qs}$ inactive and one active (see eq. (2.5)). For this last case, we have to compute the probability of detecting $u_{qs} - \beta_{qs}$ individuals, *and* having one of them active *and* all the other inactive. To get this we have to multiply by the probability $p^{act}(i')$ of having a neighbour i' active and all the probabilities $(1 - p^{act})$ to ensure all the neighbours are inactive. In addition to that we sum over i' since the active neighbour could be any of the $u_{qs} - \beta_{qs}$ detected,

$$\sum_{i=u_{qs}-\beta_{qs}+1}^{n} p_b(j, i) \sum_{i'=0}^{i} p^{act}(i') \prod_{\ell \neq i'}^{n-1} \left(1 - p^{act}(\ell)\right), \tag{2.29}$$

where the subindex i' runs for all the possible neighbours of j. We remark that the sum in i' is because the individual wants to detect i particles and have neighbour «0» active, or detect i particles and have neighbour «1» active, etc. Note that each one of the neighbour has its own activation probability. Sum over i is performed to give the possibility to detect more than $u_{qs} - \beta_{qs}$ inactive particles in the system, as in (2.28). To make the notation clearer, we define

$$p^{act}(i_1, \ldots, i_m) = p^{act}(i_1) \cdot \ldots \cdot p^{act}(i_m) \prod_{\ell \neq \{i_1, i_2, \ldots i_m\}}^{n-m} \left(1 - p^{act}(j_1)\right), \tag{2.30}$$

as the complete probability of having exactly m active particles in the system. However, we can have two active individuals and then, the individual need only to detect $u_{qs} - 2\beta_{qs}$ particles, plus 2 active, so we have the contribution

$$\sum_{i=u_{qs}-2\beta_{qs}+2}^{n} p_b(j, i) \cdot \sum_{i_1=0}^{i-1} \sum_{i_2=i_1+1}^{i} p^{act}(i_1, i_2). \tag{2.31}$$

Here one has to be careful with the limits of the sums. $p^{act}(i_1, i_2)$ is the probability of having the individuals i_1 and i_2 active, so we have to sum all the combinations with $i_1 \neq i_2$. A way to do this is to sum over i_1 to have all the possible particles in i_1, and then add a sum $i_2 = i_1 + 1$ to have all the possible pairs. If there are n particles in the system ,the last pair is $p^{act}(n-1, n)$. This implies that the sum in i_1 has to finish in $i - 1$. This procedure can be readily extended to cases with more active particles. We can apply the same reasoning until the point in which individuals can activate only by the effect of other active bacteria. The minimum number of active bacteria that we need for this is u_{qs}/β_{qs}. In this case, the contribution is

$$\sum_{i=0}^{n} \sum_{i'=u_{qs}/\beta_{qs}}^{n-i} p_b(j, i + i') \cdot \sum_{j_1=0}^{i'-n+1} \sum_{j_2=j_1+1}^{i'-n+2} \cdots \sum_{j_{i'}=j_{i'-1}+1}^{i'} p^{act}(j_1, j_2, \ldots j_{i'}), \tag{2.32}$$

where now the sum on i goes for the inactive bacteria and the one in i' for the active ones. In this case we can have $i = 0$ since the contribution of the bacteria particles is enough to activate. The sums in j are similar to the ones we have seen for only two particles. Then, the final general expression for the activation probability is given by the sum of all the terms,

$$\begin{aligned} p^{act}(j) = \prod_{\ell=0}^{n}\left(1-p^{act}(\ell)\right)\cdot\sum_{i=u_{qs}}^{n} p_b(j,i) + \sum_{i=u_{qs}-\beta_{qs}+1}^{n} p_b(j,i)\sum_{i_1=0}^{i} p^{act}(i_1) + \\ + \sum_{i=u_{qs}-2\beta_{qs}+2}^{n} p_b(j,i)\sum_{i_1=0}^{i-1}\sum_{i_2=i_1+1}^{i} p^{act}(i_1,i_2) + \ldots + \\ + \sum_{i'=u_{qs}/\beta_{qs}}^{n}\sum_{i=0}^{n-i'} p_b\left(j,i+i'\right)\sum_{i_1=0}^{i'-n+1}\sum_{i_2=i_1+1}^{i'-n+2}\cdots\sum_{i_{i'}=i_{i'-1}+1}^{i'} p^{act}(i_1,i_2,\ldots i_{i'}). \end{aligned} \tag{2.33}$$

There are two important things to remark at this point. On the one hand, in the first term we have included the probability that every cell is inactive. On the other hand, it is important to realize that even in a homogeneous distribution not all addends contribute in the same amount. This is because if we throw points at random in the space, the expected number of points detected inside a circle of radius r is given by $np = n\pi(r/L)^2$. The most important contributions to the probability of activation will come from terms around this mean value. In addition to that, in a homogeneous space the probability to be active is easier to compute. Since all the particles are equal, they have the same probability to become active and once the first ones become active, the activation propagates very fast over the whole system. Therefore, we have that p^{act} is similar to a unit step function. The point where the activation happens for the whole system can be obtained from the contribution of inactive particles (2.28), using a constant $p(j,i) \equiv p$. The form of p^{act} in a well-mixed population has an analytical solution that fulfills that $p^{act} \to 1$ as n grows, as we would expect.

A spatial analysis is more complicated, spatial structure may emerge, and each individual has each own activation probability. In any case, equation (2.33) holds in general for any distribution of particles. The value of $p^{act}(j)$ depends on the probability of detecting neighbours given that j is located at $\vec{x}_j$, as well as on the activation probability of these neighbours. Equation (2.33) is difficult to solve since it consist on a large number of coupled non-linear equations. One approximation that can be made is to suppose that the radius of detection r_{qs} is small, so the local density of particles around j is going to be approximately constant, and all the activation probabilities of the neighbours are going to be locally equal. Such approximation reads

$$p^{act}(i_1,\ldots i_m) \simeq \left(p^{act}(j)\right)^m\left(1-p^{act}(j)\right)^{n-m}, \tag{2.34}$$

and it allows us to decouple all the equations. With this approximation, the activation probabilities are constant with respect to the sums, and they can be performed analytically. In fact, by direct evaluation one can demonstrate that

$$\sum_{i_1=0}^{n-k}\sum_{i_2=i_1+1}^{n-k+1}\cdots\sum_{i_k=i_{i-1}+1}^{n} 1 = \frac{n(n-1)\ldots(n-k+1)}{n!} = \binom{n}{k}. \tag{2.35}$$

This result, together with (2.34), leads to the following expression for the activation probability,

$$p^{act}(j) = \sum_{a=0}^{u_{qs}/\beta_{qs}-1}\sum_{m=u_{qs}-a(\beta-1)}^{n} p_b(j,m)\,p_b^{act}(m,a) + \sum_{a=u_{qs}/\beta_{qs}}^{n}\sum_{m=a}^{n-a} p_b(j,m)\,p_b^{act}(m,a), \tag{2.36}$$

where now the sum in a accounts for the number of active bacteria, and m is the total number of bacteria detected. These are the only two indices needed since all the neighbours have now the same probability to activate. In addition to that, we have defined the probability,

$$p_b^{act}(m,a) = \binom{m}{a}\left(p^{act}(j)\right)^m\left(1-p^{act}(j)\right)^{m-a}, \tag{2.37}$$

which is a binomial probability. It is natural to obtain this expression, since $p_b^{act}(m,a)$ is the probability of getting a active particles from a total of m with a fixed probability $p^{act}(j)$, which is by definition a binomial. Even assuming that r_{qs} is small, the system of equations is a polynomial in p^{act} that cannot be solved analytically. In any case, the approximated equation can be solved numerically since $p^{act}(j)$ is written only in terms of $p^{act}(j)$, whereas the solution of the entire system of equations would be computationally very expensive, since we can have of the order of 10^4 coupled non-linear equations. When we solve the mean-field PDE for the density numerically we have to solve equation (2.36), in every iteration, which makes the problem computationally heavy. The approximation (2.36) is in agreement with the simulations performed for well-mixed systems, as can be seen in Figure 2.3. This check is particularly illustrative since in well mixed systems, $p^{act}(j) = p^{act}$ is a constant for every j. As we have discussed, p^{act} has approximately the shape of a unit step function in this case.

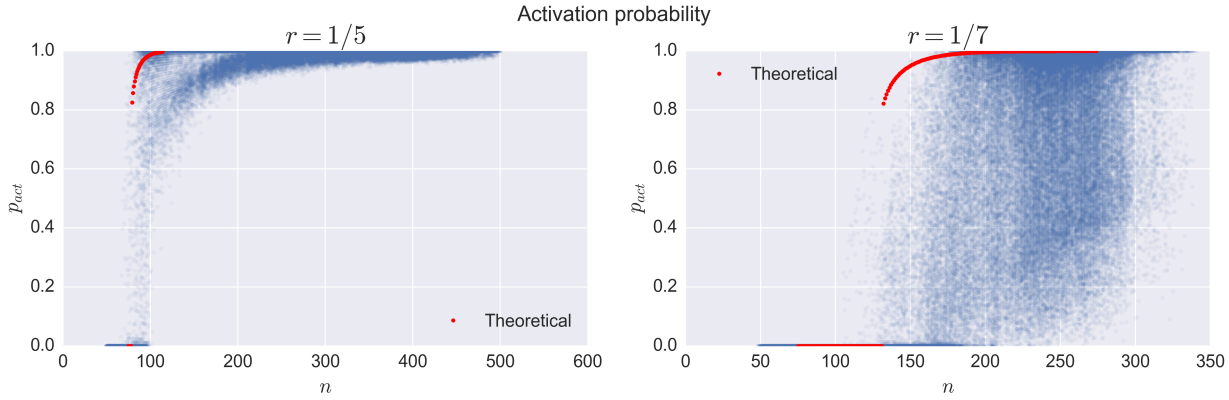

Figure 2.3: Activation probability for two different values of the detection radius r_{qs}. We show points in space from 50 different realizations. Transparency allows us to see most probable values and fluctuations. Red points show the numerical solution of p^{act} obtained using (2.36). Note its similarity with a step function. This is compatible with the values obtained for the threshold of activation u_{qs}. For lower r_{qs}, we obtain a worse agreement since the fluctuations increase. Simulations were done with $N = 600$ particles.

Now that we have characterized the activation probability, we derive the mean field equation for the density. The expected number of particles detected by individual j depends on its position $\vec{x}_j$, but it has the same definition as in the homogeneous case (2.6). This is because we define the field of detected PGs, $n^{pg}(\vec{x})$, and then identify $n_j^{pg} = n^{pg}(\vec{x} = \vec{x}_j)$. In the continuum limit we work with densities, so we need the expression for the field in any point, which is analogous to the homogeneous case

$$n^{pg}(\vec{x}) = n^{act}(\vec{x})\, p(\vec{x}, n)\, \theta\left(u_{sat} - n^{act}(\vec{x})\right) + u_{sat}\theta\left(n^{act}(\vec{x}) - u_{sat}\right), \tag{2.38}$$

and this is the only term we have to compute averages over. Assuming, as before, the mean field approximation, we get a very similar equation for $\langle n(\vec{x},t)\rangle$. Dividing by the volume, we get the expression for the densities. The mean-field equation of the complete system with space is,

$$\begin{aligned}\frac{\partial \rho(\vec{x},t)}{\partial t} &= D\nabla^2\rho(\vec{x},t) + \rho(\vec{x},t)\left[\left(1 - \left\langle p^{act}(\vec{x},t)\right\rangle\right)(k_0 - q) + \right. \\ &\left. + \left\langle p^{act}(\vec{x},t)\right\rangle(k_0 - \beta q) + \alpha u_{sat}\theta\left(\rho(\vec{x},t)\left\langle p^{act}\right\rangle - U_{sat}\right)\right] \times \\ &\times \left[1 - \frac{\delta - \alpha\left\langle P(\vec{x},t)\right\rangle\theta\left(U_{sat} - \rho(\vec{x},t)\left\langle p^{act}\right\rangle\right)}{\left(1 - \left\langle p^{act}(\vec{x},t)\right\rangle\right)(k_0 - q) + \left\langle p^{act}(\vec{x},t)\right\rangle(k_0 - \beta q) + \alpha u_{sat}\theta\left(\rho(\vec{x},t)\left\langle p^{act}\right\rangle - u_{sat}\right)}\rho(\vec{x},t)\right],\end{aligned} \tag{2.39}$$

that cannot be divided in a piecewise system, because the thresholds are local. The argument inside the step functions has been converted also to densities, defining $U_{sat} = u_{sat}/V$. Note that the effective capacity of the system the continuous version of the average growth rate of all the discrete particles divided by the death rate,

$$\rho_m(\vec{x},t) = \frac{N(\vec{x},t)}{V} = \frac{\sum_{j=1}^{n} k_j}{\delta}. \tag{2.40}$$

It is important to remark that we have also added the diffusion term that comes from the Brownian motion of the particles. Equation (2.39) is in fact a variation of a classical reaction-diffusion problem (1.33). As we discussed in the introduction, the solution of (1.33) is a front that propagates with certain velocity. In this case, however, the problem is more complicated. We will have a front that propagates, but when the density is large enough, bacteria will activate, making the population to grow faster, and producing a secondary front of very high density. This front will also diffuse and push the boundary. As a consequence, the outside front of (2.39) will move faster, as we show in Figure 2.4.

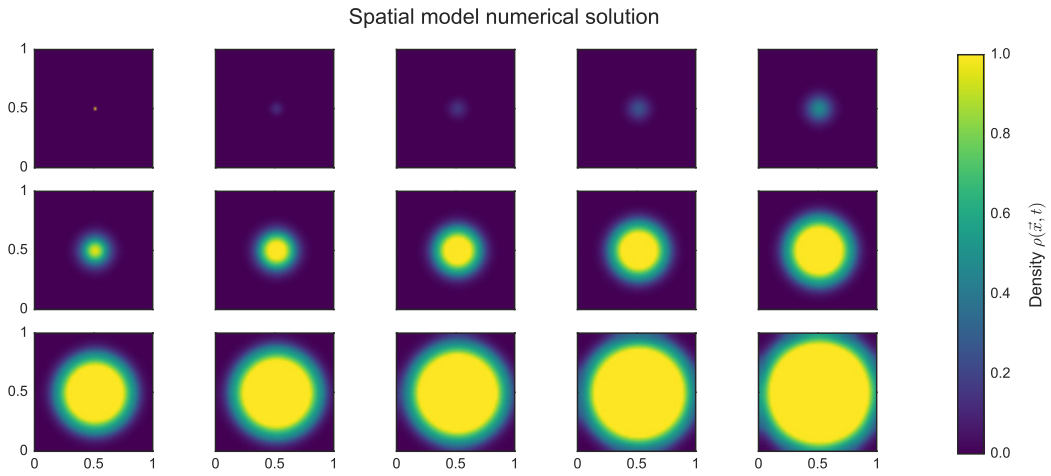

Figure 2.4: Evolution from $t_0 = 0$ to $t_f = 10$ (in arbitrary units) of (2.39) for an initial condition $\rho(\vec{x},0) = \delta(x - L/2)\,\delta(y - L/2)$. We see that it starts as the front in Figure 1.1, but as soon as the density is large enough to activate the particles, a new circular front of high density appears, increasing the velocity of propagation of the front. The outside front touches the system limit at approximately $t \sim 7$, and the inside one at $t \sim t_f = 10$.

2.3 Model parametrization.

Our model has a great number of independent parameters. Some of them are included to implicitly take into account aspects of the model, such as the radius r_{qs} and r_{pg} that accounts for the diffusion of particles, or the parameter α that controls the relative benefit of each unit of detected PG in relation to the costs. However, some others can be measured directly in experiments, such as the growth rate and metabolic costs of bacteria, k_0, q and β. In this section we do an estimation of the costs of producing AIs and PGs, using real data from experiments in the literature.

In particular, we use data from Diggle *et al.* [3] to parametrize our model. In figure 2.5, we show the growth of several *Pseudomonas aeruginosa* cultures for different bacterial strains[3]. The medium were the

[3]Originally the Y axis had an optical measure of bacterial density, called *D600*. However, for our model we need to fix a carrying capacity for numerical simulation purposes. Figure 2.5 is presented in arbitrary units fixing the maximum carrying capacity $N = 6000$

experiment is performed (called lysogeny broth or LB) has several peptides, vitamins and minerals that enhance the bacterial growth and make PG molecules useless, since bacteria can get the nutrients directly without having to decompose them. In addition, the medium is well-mixed thanks to a constant shaking at 200 r.p.m [3], so we should be able to apply our mean field theory to extract some parameters from the data. Bacteria grow until the point in which they become active. At this point, they start producing PGs. In normal conditions this enhances growth, but in this medium there is no extra benefit from the PGs, so they only have additional costs, which reduce the carrying capacity of the population.

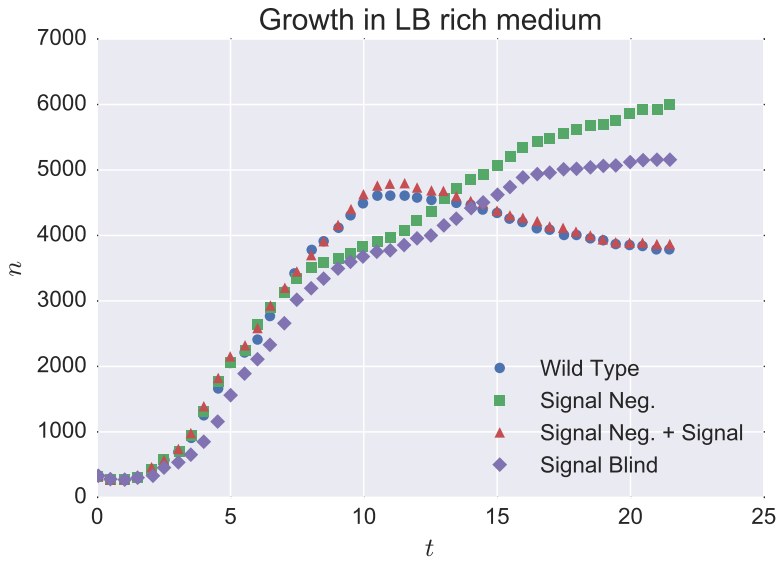

Figure 2.5: Experimental data reproduced from [3] for cultures of different strains of *Pseudomonas aeruginosa* in a LB rich medium. Note that signal negative mutants are grown with and without QS signal in order to check the costs of the PG production.

We begin the parameter estimation with the curve of the wild type strain. On the first stage of the growth, where all bacteria are inactive, the growth is logistic, with a growth rate $k_{\text{WT1}} = k_0 - q + \alpha u_{sat} \equiv k_{eff} - q$, as we saw in equation (2.21). Note that the medium has nutrients that provide a benefit to bacteria, making PGs useless. For this reason, we assume that the benefit provided by the medium is similar to the maximum gain provided by PGs, so we added the $+\alpha u_{sat}$ term that corresponds to the benefit of the medium. We can see in the data that the population decreases when bacteria activate. In the second stage of the growth of the wild type bacteria, we have $k_{\text{WT2}} = k_{eff} - \beta q$. We can fit both stages by a logistic function (1.16), and then compute $k_{\text{WT1}} - k_{\text{WT2}} = q\,(\beta - 1)$. From this we can get information about the costs of producing exoproducts. This procedure of fitting the two stages is done for every strain, since each one has different terms contributing to the growth rate, allowing us to compute more parameters, or even get different estimations of the same parameter to check if they are coherent. The results are given in table 2.1.

for the maximum measured *D600* density. However, growth rates are independent of the value of Y axis. The different values of Y axis have consequences in the carrying capacity, meaning that will give different values of δ.

Strain	Growth rate (hours^{-1})	Carrying Capacity (bacteria over 6000)	Delta$\cdot 10^{-5}$ (hours^{-1})
WT	$k_{\text{WT1}} = k_{eff} - q = (0.411 \pm 0.015)$	$N_{\text{WT1}} = (5900 \pm 500)$	$\delta_{\text{WT1}} = (6.9 \pm 0.8)$
	$k_{\text{WT2}} = k_{eff} - \beta q = (0.187 \pm 0.021)$	$N_{\text{WT2}} = (3590 \pm 50)$	$\delta_{\text{WT2}} = (5.2 \pm 0.6)$
SN+S	$k_{\text{SNS1}} = k_{eff} = (0.430 \pm 0.014)$	$N_{\text{SNS1}} = (5600 \pm 400)$	$\delta_{\text{SNS1}} = (7.6 \pm 0.7)$
	$k_{\text{SNS2}} = k_{eff} - \beta_{pg} q = (0.22 \pm 0.05)$	$N_{\text{SNS2}} = (3700 \pm 70)$	$\delta_{\text{SNS2}} = (6.1 \pm 1.5)$
SB	$k_{\text{SB}} = k_{eff} - q = (0.366 \pm 0.007)$	$N_{\text{SB}} = (5010 \pm 70)$	$\delta_{\text{SB}} = (7.3 \pm 0.2)$
SN	$k_{\text{SN}} = k_{eff} = (0.382 \pm 0.010)$	$N_{\text{SN}} = (5520 \pm 100)$	$\delta_{\text{SN}} = (6.9 \pm 0.3)$

Table 2.1: Fit of the parameters.

Once we have the fit for all the growth rates and carrying capacities, we can obtain the rest of the parameters. For example, δ can be determined for each fit as $\delta = k/N$, meaning that we have different values for δ that should be coherent. Then, we can make differences between the growth rates in order to estimate some of the costs and growth rates. The problem is that everything depends on q, that has to be determined in a different way. In the first stage of the growth curve, the wild type culture and signal negative + QS signal are identical except for the production of AIs. Then, there is a difference of q between their growth rates. We approximate the maximum of bacteria for this two strains by a parabola, and compute difference given by $N_{\text{SNS}} - N_{\text{WT}} = q/\delta$. Apart from the value of q, the point where the activation happens, u_{qs}, is the point where both the wild type and the signal negative populations reach the maximum value. Such value can be obtained from the fit of the parabola, and it is $t \simeq 11.5$ hours. Knowing the value of q we can get all the other costs. The results and the method to obtain the costs are summarized in Table 2.2. It is possible to compute u_{qs} imposing that in the whole system we expect to see one active bacterium. This can be done computing

$$np^{act}(j) = 1 \rightarrow n \sum_{i=u_{qs}}^{n} \binom{n}{i} \left[\pi \left(\frac{r_{qs}}{L}\right)^2\right]^i \left[1 - \pi \left(\frac{r_{qs}}{L}\right)^2\right]^{n-i} = 1 \tag{2.41}$$

where we use only the first term of equation (2.33), since as we discussed, $p^{act}(j) = p^{act}\ \forall j$ in the system and in this point there is no other active bacteria, meaning that the probability to detect active neighbours is zero and the other terms of (2.33) can be neglected. The activation happens very close to the point that we have computed solving equation (2.41), with an average relative error of 2%.

Parameter	How was obtained	Value	Average
k_{eff}	Directly from SN+S1	0.430 ± 0.014	0.406 ± 0.012
	Directly from SN	0.382 ± 0.010	
δ	Last column Table 2.1		6.7 ± 0.9
q	Difference between SN+S and WT at maximum	0.012 ± 0.005	0.012 ± 0.005
β	$(k_{\text{SNS1}} - k_{\text{WT2}})/q$	20 ± 12	18 ± 10
	$1 + (k_{\text{WT1}} - k_{\text{WT2}})/q$	20 ± 9	
	$(k_{\text{SN}} - k_{\text{WT2}})/q$	16 ± 9	
β_{pg}	$(k_{\text{SNS1}} - k_{\text{SNS2}})/q$	17 ± 12	15 ± 11
	$1 + (k_{\text{WT1}} - k_{\text{SNS2}})/q$	17 ± 8	
	$(k_{\text{SN}} - k_{\text{SNS2}})/q$	13 ± 10	
β_{qs}	$\beta - \beta_{pg}$	3.28 ± 0.01	3.28 ± 0.01

Table 2.2: Computation of parameters from fitted data.. Error for δ was obtained as the maximum between mean of errors and standard deviation. The difference $\beta - \beta_{pg}$ gives the same value for the three pairs of β and β_{pg} computed.

This is not an exact fit but an estimation. We assumed that the benefit provided by the medium is the

same that the maximum benefit provided by the public good. This approximation is not true, as we can see in the fact that wild type and signal negative plus signal strains have a higher growth rate than signal blind and signal negative. This happens because signal negative strain with QS signal and the wild type strain are producing small amounts of public good, controlled via QS. The PGs produced do not provide a big benefit, but the effect can be appreciated in the difference in the fitted values. However, note that all the fits are consistent.

One important detail is the fact that the cost of producing the PGs is much larger than the cost of the extra production of autoinducer. This is because the autoinducer regulates the production of PGs, which is very costly, and the cost of the regulation has to be low enough so it is worth to have a regulatory system, as we discussed in the introduction. Also, βq is smaller than the effective growth rate $k_0 + \alpha u_{sat}$, as we could expect. However, note that from the data we cannot split the contribution of the base growth rate k_0 from the contribution of the public good αu_{sat}. Therefore, we still have to decide which is the benefit of the PG in the system, that remains free. Furthermore, the point at which the saturation happens, u_{sat} cannot be estimated from the data, leaving α and u_{sat} free.

Despite our strong assumptions discussed above, the most relevant results of this estimation are the values of the costs β, which are compatible with the qualitative results we expected from biology, as well as the computation of the threshold of activation, u_{qs}. Apart from that, we have the correct order of magnitude of the parameters k and δ that make the time scale of system to be in hours. However, we have to ensure that $k_0 - \beta q + \alpha u_{sat} \geq k_0 - q$, in order to have a higher carrying capacity for the active population than for the inactive one.

3. Numerical simulations and results.

3.1 High diffusion limit.

3.1.1 Single population

In this section we perform simulations for a single population of a wild type strain, and discuss the results. We consider that the medium is in a well-mixed regime. This regime can be obtained by making the diffusion constant $D \gg 1$, so there is no spatial structure in the system.

One of the first things we want to study is the accuracy of the mean field approximation and the relation ship between this result and the detection radius. One of the interpretations of the mean field approximation is that all the individuals of the system are interacting with all the individuals, so there is no stochastic fluctuations. Therefore, it is interesting to study how the system behaves when the radius of detection changes from $r \sim L$ to $r \ll 1$, since in the first case we expect to see a behaviour very similar to the one in the mean-field approximation, while for $r \ll 1$ fluctuations become important.

We want to know how the radius of detection affects the fluctuations in order to be able to control fluctuations in the system. The number of bacteria in a Petri dish is very high, so the mean field approximation describes well their behaviour because the fluctuations decrease as $N^{-1/2}$. However, we cannot simulate such number of bacteria. We need a number that we can simulate in a reasonable time, but high enough so fluctuations can be neglected to apply the mean field approximation. In Figure (3.1), we show the effect of changing the interaction radius $r = r_{pg} = r_{qs}$ on the average growth rate in a system with fixed carrying capacity. Recalling that in a system where all bacteria are active we have

$$\langle k(n) \rangle = k_0 - \beta q + \alpha \langle n \rangle , \tag{3.1}$$

we expect that the growth rate increases in a linear way up to the saturation point. We observe such linear growth for large values of the radius. However, in Figure 3.1 we can see that for $r = L/6$, the fluctuations are very large and the behaviour is different. In fact, when fluctuations are so large the mean field description is not good and the expressions of $\langle k(n) \rangle$ are not correct. The effect of the fluctuations usually gives a decrease of the growth rate: individuals that detect fewer active neighbours than the average get a lower benefit, but individuals that detect more neighbours than the average are not getting a big benefit since the detected PGs are limited by u_{sat}. The reduction of the average growth rate translates to a lower carrying capacity of the system. For this reason, we decided to increase the carrying capacity of the system up to the order of $N \gtrsim 5000$ to perform the simulations. In this regime, for $r_{pg} = 1/6$ fluctuations are low and our mean field description works.

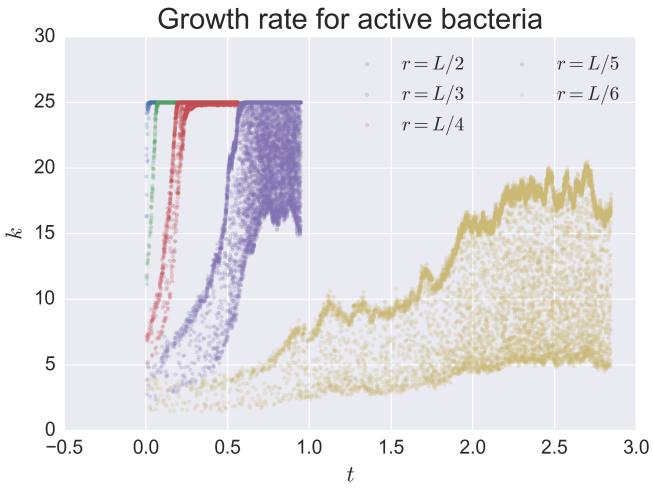

Figure 3.1: Evolution of the average growth rate in the system for different interaction radius and fixed $N = 600$ particles. Transparency is used to illustrate fluctuations. Notice that curves whose area of interaction is similar to the system size follow a straight line and then a saturation, with no dispersion of points, as we expect in mean field. For $r \leq L/6$, fluctuations are very important and the mean field computations are no longer valid.

In Figure 3.2, we show the comparison between the theoretical mean field equations and a stochastic realization of the individual model in a well-mixed regime. Both results are very similar, and allow us to see that for $r_{pg} = 1/6$ and $N \geq 6000$, the mean field describes accurately the behaviour of the system, as we expected.

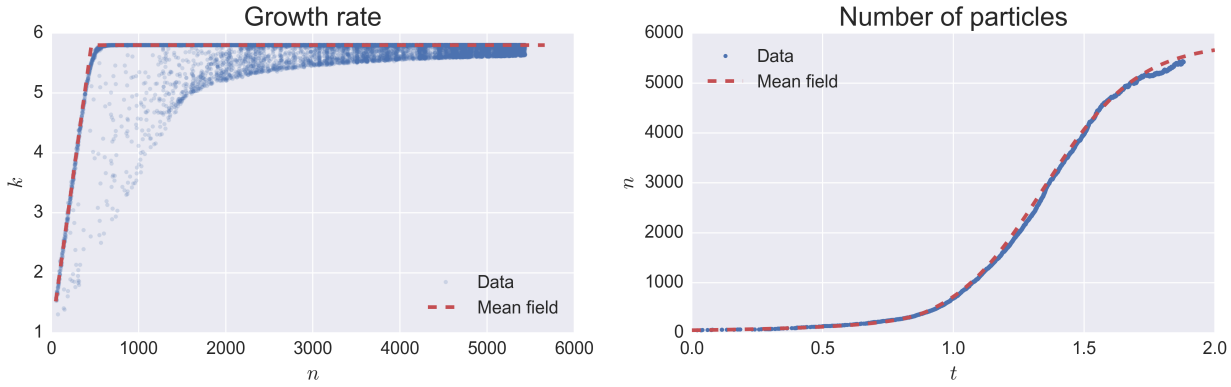

Figure 3.2: Comparison of the mean field approach and a stochastic realization with a higher carrying capacity, $N = 6000$, and interaction radius $r = r_{qs} = r_{pg} = L/6$. Note that in this case the system follows very well the mean field assumption, in contrast with Figure 3.1. This is because of the increase in the carrying capacity. Figure in left has transparency to show the strength of fluctuations.

In addition to that it is possible to perform simulations of the rich medium and compare with the experimental data by Diggle *et al.*, [3]. The LB rich medium is providing the maximum benefit possible to the bacteria, so the base growth rate is given by $k_{eff} = k_0 + \alpha u_{sat}$. As we see in Figure 3.3, the behaviour of the system is very similar to the experiments (Figure 2.5), with the difference that at the point where bacteria

become active, the curve is not smooth but it has a spike. This is because we are doing an approximation for the growth rate. It is a continuous non-linear function with a logistic behaviour [7], but we instead use a linear function with a saturation point. For this reason, we have a discontinuous jump between the rate $k_{eff} - q$ of the inactive stage and $k_{eff} - \beta q$ of the active one. Given that the carrying capacity decreases when the individuals activate, the population decreases until its new carrying capacity, where it remains stable. Also, note that in these simulations the timescale is measured in hours, while the units of the Y-axis are an arbitrary carrying capacity that can be converted to bacterial density, meaning that it is possible to compare results from the simulations directly with experiments.

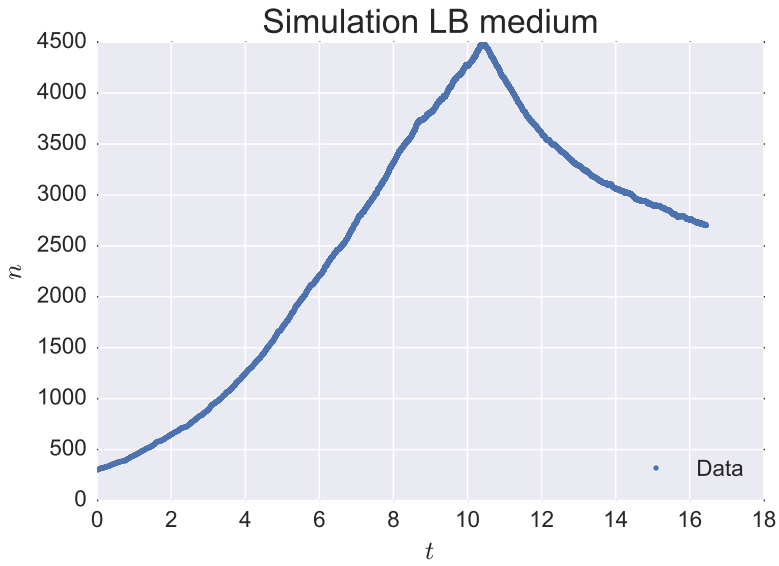

Figure 3.3: Simulation of the wild type strain for the parameters estimated in Section 2.3. We can see that the activation of the system happens at $t \sim 10$ hours, very similar to the value obtained by Diggle *et al*. in the experiments. In addition to that the value of the population at the maximum is very similar to the one shown in our arbitrary scale in Figure 2.5, reproducing the experiments accurately.

3.1.2 Mixed populations

In the case of mixed populations, we want to address the problem of how non-producing strains can take profit of the PG producers in a population. In our model, the external good is public in the sense that it is available for all the other cells in the system [5]. For this reason, it can be exploited by non-producers [4, 5]. As we discussed in Section 1.1, we have two kinds of cheater mutants that are able to exploit the PG production: signal negative (does not produce autoinducer signal) and signal blind (does not react to signal). The signal blind strain, in particular, is the non-producer mutant, while the signal negative strain still produces PGs, but not AIs. The advantage of the signal negative strain over the wild type cells comes from the fact that it does not produce the autoinducer signal. In this section, we study the evolution of mixed cultures without spatial structure. To quantify the relative growth between the wild type and the cheaters, we compute the relative fitness of the cheaters,

$$f = \frac{w_2\,(1 - w_1)}{w_1\,(1 - w_2)}, \tag{3.2}$$

where w_1 and w_2 are the initial and the final fraction of cheaters in the system. We will have that $f > 1$ if the cheater strain grows more than the wild type. Then, we explore how the relative fitness f depends on

the initial fraction of cheaters, w_1. As we discussed in the introduction, when w_1 increases, the amount of PG available in the system decreases due to the lower density of wild type bacteria. The consequence is a lower growth rate and carrying capacity. The growth of the cheater directly depends on the PGs produced by the wild type, so the decrease in PGs will translate into a lower relative fitness. This has been checked in experiments, such as [3]. Here we perform simulations of mixed cultures in order to compare these experimental results with our model.

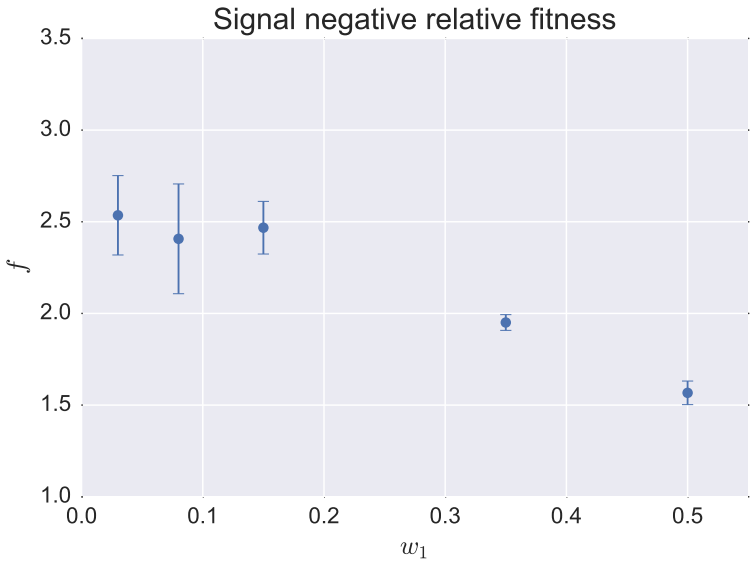

Figure 3.4: Relative fitness of the cheater for a mixed population of wild type and signal negative, averaged over 10 realizations. We can see that the relative fitness decreases as we increase the initial number of cheaters, showing the negative frequency dependence. When the initial number of signal negative is high enough, collective activation is not triggered, PG is not produced, and we have a high decrease of the relative fitness produced by a lower carrying capacity. Parameters for simulation were $k_0 = 0.35$, $\delta = 6.95 \cdot 10^{-5}$, $\alpha = 10^{-3}$, $q = 0.01196$, $\beta = 22.08$, $u_{qs} = 300$, $u_{sat} = 470$. The number of Gillespie iterations was $8 \cdot 10^4$. Note that the cost's value come directly from the estimation we did in Section 2.3. This set of parameters is used in all the following simulations.

First, we discuss the behaviour of a mixed culture of signal negative and wild type bacteria. The signal negative mutant does not produce the autoinducer molecule, but it is still able to become active in the presence of the signal and it produces public good. The signal negative mutant only affects the communication channel, so it is qualitatively different from a non-producer cheater as the signal blind. Both the wild type and the signal negative strains need the signal molecule in order to activate and start producing PG. When we have initially a low amount of signal negative cells, the wild type grows and produces AIs until both strains activate and PG production starts. Once this happens, the maximum carrying capacity is the same as in the single-species population, since PG can be produced by both strains. The signal negative strain does not produce signal, so its growth rate is higher. As a consequence, the proportion of wild type cells will decrease (see Appendix B). On the other hand, when the initial proportion of the cheater is high, the system will saturate before starting the activation phase. In this case the PGs will not be produced and the carrying capacity will reduce dramatically, reducing even more the relative fitness of the cheaters. This effect is known as the *tragedy of the commons*, applied to the communication level. The dependence of the relative fitness f on the initial proportion of cheaters w_1 can be seen in Figure 3.4.

The study of the invasion of a population by non-producers, i.e. bacteria that does not produce PG, is

a classical problem in the context of social evolution [4–6, 23]. The signal blind mutant does not react to the autoinducer signal, so it is a non-producer mutant. The point where wild type bacteria activates and starts producing the PG is the same as in a single-species wild type population, since the signal blind still produces signal at the same rate as the wild type. Then, we will have production of PG by wild type, which will increase the carrying capacity until certain limit, which is controlled by the initial fraction of cheaters. This is because the fraction of signal blind bacteria at the activation point is proportional to initial amount of cheaters. The final carrying capacity of the system will depend on the stationary number of wild type bacteria, which is lower as the number of cheaters increase. The consequence is that the relative fitness decreases as the initial number of signal blind cells increases. In Figure 3.5 we can see the relative fitness f of the signal blind cheater when we increase its initial proportion. It is worth to mention that in these simulations, the wild type population was able to resist the invasion of the non-producer invaders, making $f < 1$. In real experiments, such as [3], the signal blind strain grows in the system with $f > 1$. We do not get the same quantitative results due to our election of parameters for the system. The need for tuning the parameters in order to have a quantitatively correct relative fitness for the cheater is a disadvantage of the model. However, we remark that the important result of Figure 3.5 is the reduction of the relative fitness as w_1 increase. The particular values of the Y axis has to be tuned correctly with the adequate parameters.

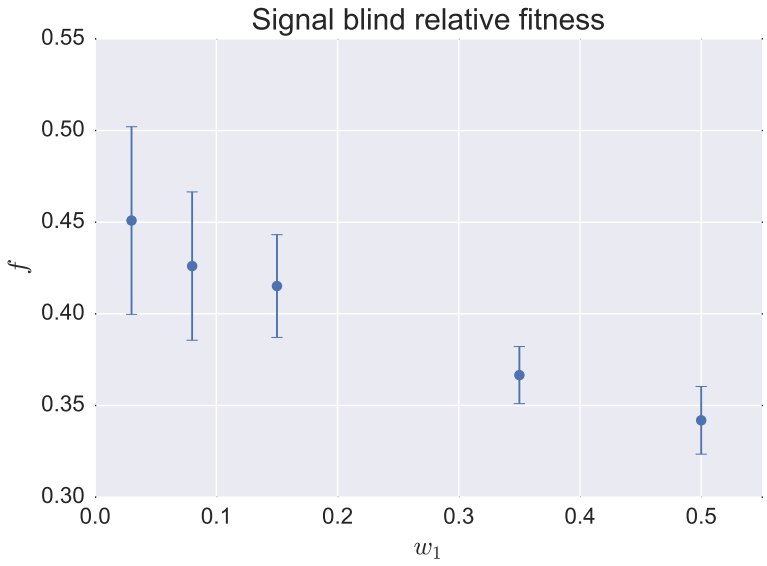

Figure 3.5: Relative fitness of the cheater for a mixed population of wild type and signal blind. We can see that the relative fitness decreases as we increase the initial number of cheaters. However, quantitative values of the fitness do not coincide with experiments.

The difference between the two strains is clear. The signal blind mutant shows a continuous decrease of f as we increase the initial number of cheaters, while the signal negative has a faster change for a certain threshold of cheaters, the one that avoids cell activation in the system. Both cheaters should be able to outcompete the wild type strain, but the model needs a tuning in order to achieve this. Depending on the growth conditions and parameters, both populations can coexist, or one of the strains can extinct. In order to have coexistence, f should have a point such that $f(w_1) = 1$ for some w_1. At this point, both strains grow at the same rate, leading to a coexistence of the strains. Note that for our parameters, signal negative have $f > 1$ for all w_1, while signal blind has $f < 1$ for all w_1, leading always to the extinction of one of the strains in all the cases.

3.2 Low diffusion. The role of space.

3.2.1 Single population.

Until now, we have studied simulations for a well-mixed population, with $D \gg 1$. In this final section, we wonder what happens when the system has some kind of spatial structure, both for single and mixed cultures. In the case of mixed cultures, some authors have shown that the spatial structure can be an efficient strategy to limit the effects of cheating [4, 23], so we want to analyse our model to see whether these results hold. Moreover, until now all the simulations assume $r_{qs} = r_{pg}$. Here we wonder what happens when we use a more realistic assumption, $r_{qs} \geq r_{pg}$, based in the fact that the diffusivity of the AI is greater than the diffusivity of the PGs. Basically we want to study which results change when we include a spatial structure. In addition to that, we wonder whether the results are compatible with experiments results.

We start by checking what is the role of space in the growth of a single population. We first study the differences in growth between the well-mixed regime and a cluster of bacteria that grows. Both very high values of D (that correspond to a well-mixed system) and very low values of D in a cluster (where every particle can detect the whole system) can be identified with the mean field approximation. We are interested in an intermediate regime where the diffusion is not able to destroy the spatial structure but allows bacteria to move around the system. We decide to study the range of values $D = [0.01, 0.001]$. For $D = 0.01$, $\langle x^2 \rangle \sim L^2$ for $t \sim 50$ h, which is larger than the timescale we work in our simulations. As we can see in Figure 3.6, a small diffusion constant is able to make the system grow faster than the well-mixed regime, since low diffusion enhances the clustering of particles, that will lead to a fast activation of the public good production, as well as an increase of the PG detected.

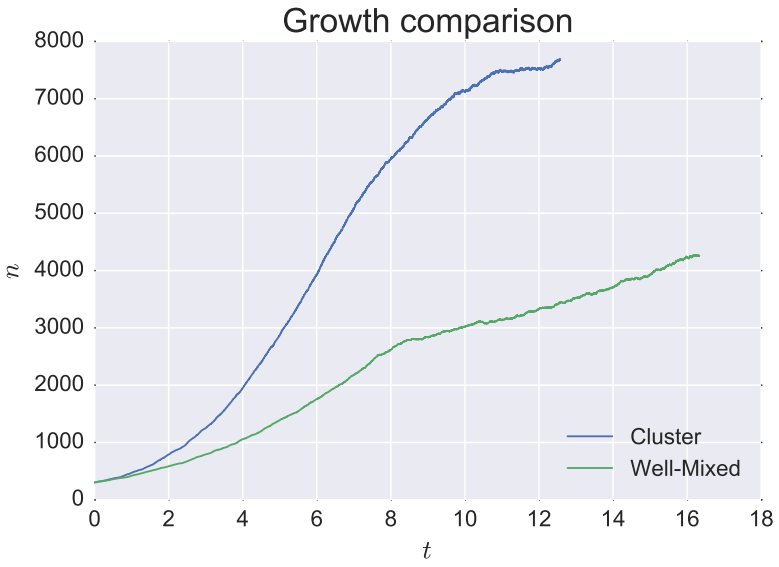

Figure 3.6: Growth of wild type bacteria in a cluster with $D = 10^{-3}$ and in a well mixed population. We can see that the growth of the clusted system is much faster.

Since clustering enhances the growth, it is also interesting to study how the spatial structure arises from random initial conditions. This has been studied for birth and death processes with diffusion, which lead to cluster formation in the system [24]. The particles form clusters since death can affect any particle in the system, while new particles coming from births are placed next to its progenitor. The quorum sensing mechanism also plays a role in this structuring, since a clustered region has a higher density of particles,

which will activate. Once cells become active, they reproduce faster, making the cluster grow. In Figure 3.7 we can see how a non-homogeneous spatial distribution can appear from initial random conditions.

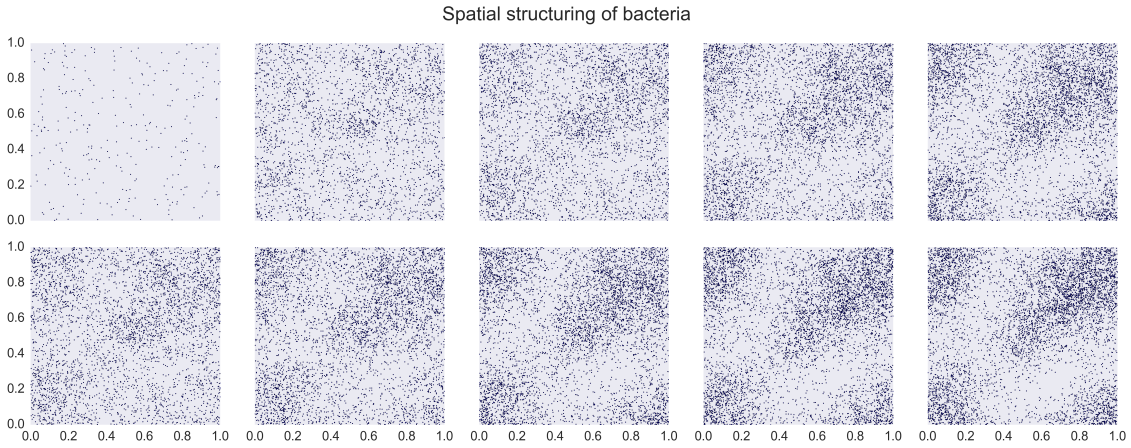

Figure 3.7: Formation of spatial structures due to the clustering and quorum sensing mechanism. We start with 300 particles and random initial conditions. At the end, we have near 8000 particles forming a cluster. Periodic boundary conditions are used.

Finally, it is interesting to study the growth of a single cluster, in order to compare with the result obtained in Figure 2.4. We simulate an initial number of particles situated at the center of the system, and observe the cluster growth. This is shown in Figure 3.8. The result is compatible with the result shown in Figure 2.4: there is a high density of particles in the center, where the growth is faster due to the activation of individuals. Bacteria diffuse from the center to the outside of the cluster, enhancing the growth.

Figure 3.8: Growth of a cluster of wild type bacteria. We see that there are two different fronts, one of high density in the center, and the one that defines the cluster itself. The population stops growing when the carrying capacity is reached.

3.2.2 Mixed populations.

In this section we study what is the effect of space in mixed cultures. The case of mixed populations is very interesting. Some studies have tackled the role of space and the structure of the population when there are non-producer individuals, both at a theoretical level [23] and experimentally [5, 6, 25]. These studies found that spatial organization can minimize the invasion of non-producers in a scenario with diffusive PGs. Moreover, Mund *et al. [25]* found that in a system regulated by quorum sensing the autoinducer molecule can act as a public good itself. To prove it, they performed experiments changing the diffusion constant of the signal molecule, and showed that when the autoinducer becomes more «private» to the producers, it helps against the invasion of cheaters in the system.

First, we study a mixed culture of wild type producers and non-producers signal blind bacteria. We fix all the parameters of the system, using an initial fraction of cheater of $w_1 = 0.1$, and change D between 0.001 and 0.01. In figure 3.9 we show the relative fitness of the cheater as a function of the diffusion of bacteria in system. As we expected, the relative fitness of the cheaters is reduced as D decreases, i.e., as the spatial structure (clustering) becomes more relevant. This is because birth is a local process, so a system with births and deaths and local diffusion will tend to form clusters [24]. Cheaters will grow faster and form a cluster. Since a non-producer cannot obtain benefits from other non-producers, the individuals in the center of this cluster will not reproduce that fast. On the other hand, the wild type strain will also form clusters in space. If the PGs do not diffuse very far, the benefits of PG will be only for the cluster of producers. There is an interplay between the diffusion and the radius r_{pg} which controls if the external good is public or private. Figure 3.9 is coherent with results of [5, 6, 23], except that the signal blind cheater should outcompete the wild type in most of the cases. As we discussed before, this is a problem of model parametrization.

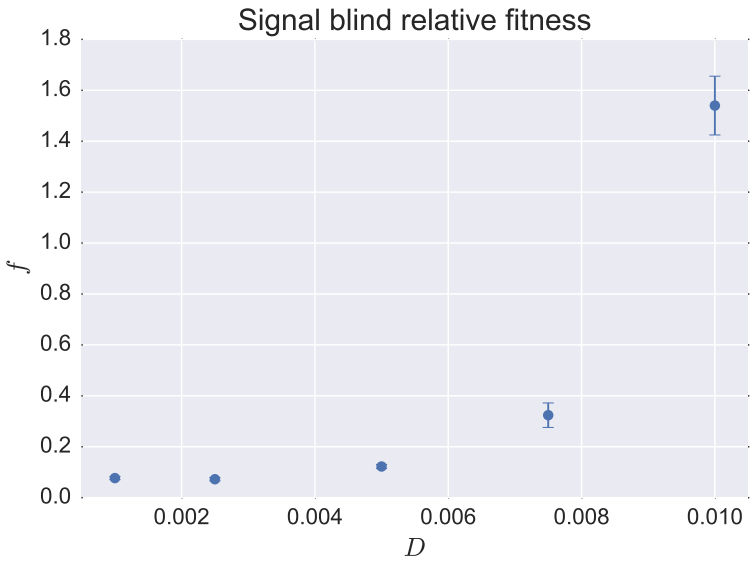

Figure 3.9: Relative fitness for mixed culture of signal blind and wild type, for an initial number of cheaters of 10% of the system. For $D \to +\infty$ we should recover the values of Figure 3.5. This simulation was done for the same parameters, but $5 \cdot 10^4$ Gillespie steps. The value of the signal blind fitness at the same time for this concentration, in the well mixed limit is approximately $f \sim 2.0$. We can see that going for lower D (more structured space) the relative fitness decreases.

Finally we want to see the dependence of the interaction scales on the relative fitness of the cheaters. We fix the diffusion coefficient $D = 10^{-3}$ so spatial clusters are promoted, and study the dependence of f with the distance r_{pg} for a mixed culture of signal blind and wild type strains. Results can be seen in the left plot

of Figure 3.10. The relative fitness of the signal blind bacteria increases when r_{pg} increase. This is because an external good available for all the particles in the system can be exploited by the cheaters, that increase their relative fitness. Producers can minimize the invasion of cheaters by structuring in space in such a way that the PG does not reach the cheaters. This is easier to accomplish if r_{pg} is small.

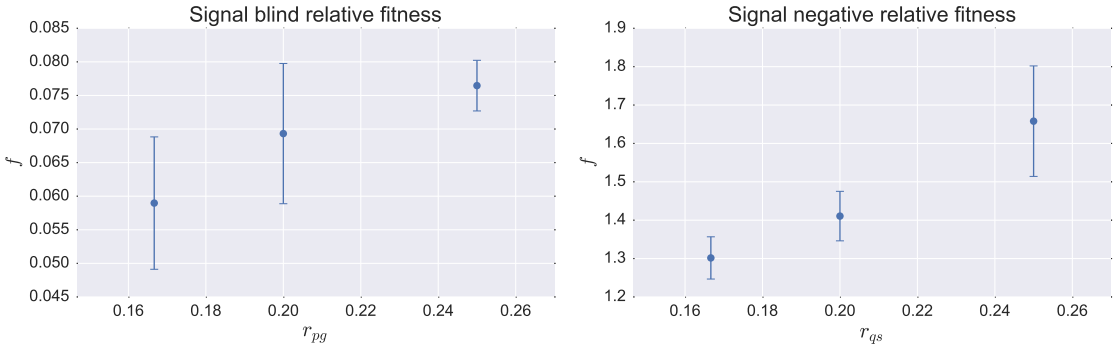

Figure 3.10: Effect of changing the interaction scales $r_{pg}$and r_{qs} for the signal blind and signal negative strains, respectively. We can see that increasing the interaction distance (making the exoproduct more «public») increases the relative fitness of the cheaters. Initial concentration of cheaters was 10% for signal blind and 60% for signal negative.

Finally, we check what happens when we fix r_{pg}, but we change the diffusion of the signal (mediated by r_{qs}), in a mixed culture of wild type and signal negative strains. As we discussed before, the benefit obtained by this mutant depends on the communication channel, so we wonder what happens when we make more «public» the signal molecule. We observe a very similar effect: when the AI molecule becomes more public for the system, the signal negative strain increases its fitness. This can be seen in the right plot of Figure 3.10. Note that for this simulation we need a relatively high number of cheaters, a proportion $w_1 = 0.6$, since for low initial densities, they are not able to avoid the activation of the wild type. To see the relative fitness decrease for an initial low density of signal negative, we need to have a very low interaction radius, which leads to high statistical fluctuations. It is important to note that Mund *et al.,* in [25], found the same qualitative results in an experimental setup. This finding lead them to claim that the signal molecule is itself a public good, since it can reinforce the growth of the population, and making it more private helps the wild type to resist an invasion of signal negative cheaters.

Note that we did the simulations for the PGs with signal blind cheaters and the simulations for AIs with the signal negative mutant, due to the different nature of these strains. We remark here that the signal blind is a non-producer, so it is interesting to see how the diffusion of PGs affects their growth. On the other hand, the signal negative strain affects only the communication channel, making it suitable to study its relation with the diffusion of the signal.

4. Conclusions and future work.

In this work we have presented and analysed an individual based stochastic model for a public good production regulated by quorum sensing. We have analysed analytically the case of a well-mixed population of a single species, and performed simulations to study the effect of a cheater population in the system and the role of the space in the outcome of the interaction.

The study of the well mixed model for a single species case allowed us to get the mean field equations of the system, as well as to check the numerical simulations. The growth of such species is a modified logistic growth, that has three different stages: inactive individuals below QS threshold, production of PG due to above the QS threshold, and saturation of the PG detection. These three stages are in fact logistic functions, with different carrying capacities and growth rates. As a result, the growth process is also logistic-like. Then, using the mean field analytical results we could estimate the parameters of the system, using experimental data from Diggle *et al.* [3]. Although we could not get all the parameters, it was possible to determine all the metabolic costs of AI and PG production. In this way we found that the extra production of the autoinducer molecule increases by three times when bacteria activate ($\beta_{qs} \simeq 3$), while the PG production has a cost that is approximately 20 times larger than the cost of AI production of an inactive individual ($\beta_{pg} \simeq 20$). These results are compatible with the fact that quorum sensing is used to control a costly production of public goods. The production of PGs at low bacterial densities would reduce the fitness of the population due to the high cost of PGs. The QS mechanism allow the system to avoid this high cost until the density of the system is large enough, so the PGs will provide a benefit for the entire population. It was not possible to estimate the benefit provided by the PG, meaning that the values of k_0, α and u_{sat} were freely chosen, considering the fact that we should have $\delta - \alpha \langle P \rangle < 0$ in order to have the exponential growth while $u_{qs} < n < u_{sat}$.

The next step was to study the effect of a cheating mutant in the system. Diggle *et al. [3]* found that increasing the initial proportion of a signal blind cheater in the system, their relative fitness decreased. This result is known as the tragedy of the commons effect: when the quantity of non cooperators increases, its growth rate decreases because the population of cooperators decreases, leading to lower amount of PGs in the system. We performed simulations for two different mixes: (i) a wild type and a signal blind mutant (ii) a wild type and a signal negative mutation. In the case of signal blind strains, we found the same qualitative results as [3]. In addition to that, we showed that signal negative strain works in a qualitative different way. The difference between the two mutants is due to the fact that the signal blind mutant is the classical non-producer strain, so it takes profit of the PG generated by the wild type, while the signal negative strain can turn down the activation of the population, avoiding the production of PGs to start. However, we remark that both strains have a negative frequency dependence of the relative fitness on the initial fraction of cheaters, so both are affected by the tragedy of the commons.

Finally, we studied the role of the space on the results. We saw the difference in the growth of a wild type population due to the spatial structure, and determined that clusters enhance the growth of the system. When a signal blind cheater is added, we wanted to study whether structure in space is able to minimize the effects of the cheater, as shown in [4,5,23]. It turned out that reducing the diffusion constant of bacteria, the relative fitness of the cheater is reduced, in accordance with this claim. Moreover, Mund *et al.* [25] found that the signal molecule can be considered itself as a PG. Its diffusivity can affect the growth of the population, depending on the spatial structure. We also checked this, reducing r_{qs} for a fixed value of r_{pg}, finding that if

the exoproducts are more public, then the cheaters can increase its fitness. In the case of the signal molecule, this is the same result found in experiments by Mund *et al.* in [25], so we can say that even the AIs can act as public goods themselves.

Even when the model is able to reproduce some results, it has drawbacks that we discuss here. The main advantage of individual based models is that they are relatively easy to simulate, and they are useful to relate microscopic and macroscopic variables. Identifying relevant biological parameters becomes easier, since we can do it from the microscopic perspective, where we have knowledge about what an individual bacterium does. In contrast, the problem of an individual based model is that the complete macroscopic equations, such as (2.39), are very complicated to solve, even numerically. For instance, to solve equation (2.39), we had to do important approximations on the value of the activation probability $\left\langle p^{act}\left(\vec{x},t\right)\right\rangle$, which is one of the fundamental variables for the macroscopic description of the quorum sensing mechanism.

Our model has been successful in explaining some experimental results. Even if the model was motivated by *Pseudomonas aeruginosa*, and we focused on replicating results from Diggle *et al.* [3] about the interaction between wild type cooperative and non-cooperative mutant strains, our results are consistent with more general results obtained in the context of the evolution of social behavior, such as the role of the spatial structure in evolution of the cooperation [4, 5, 23], the differences between «public» and «private» goods [5] or the effect of the diffusivity of the AIs in the system [25]. The success of the model explaining different results indicates that it is able to capture the most important features of quorum sensing and PGs production. On the other hand, the model requires some information about the microscopic individuals that we could not obtain. In Section (2.3), we did an estimation of the parameters for the model based in the data from Diggle *et al.* [3]. However, since the data were referred to a medium in which bacteria grow with no benefit from PGs, it was impossible to determine the benefit provided by the PGs. Nevertheless, even if we could obtain this benefit, from experimental data it is difficult to obtain the concentration of PGs u_{sat} for which the individual does not increase its benefit when receiving more PGs, so we cannot do an estimation for parameters α or u_{sat} separately. We had additional problems trying to tune correctly the free parameters, that resulted in an unrealistic behaviour for the relative fitness of the signal blind cheater. The model gives qualitative correct results for the dependence of the relative fitness of the single blind strain, but the quantitative results disagree with experimental data.

In addition to that, it is also difficult to relate the variables r_{qs} and r_{pg} with experiments. These two variables are used to model the diffusivity of AIs and PGs implicitly, and the detection of these molecules is done via the detection of neighbouring cells. This means that we can not account for the behaviour of the concentration of the exoproducts. This is important, since the diffusivity of PG molecules has attracted a lot of attention in the context of evolution of cooperation [4–6,23]. The model has demonstrated that the implicit representation of the exoproducts is enough in order to recover qualitatively some of the main results of those studies, but there is no chance of relating the radius of detection with real experimental data. We have considered the diffusion of this molecules implicitly as a first approach that saves the computational cost of modeling both substances. Building a more complete model that also includes PGs and AIs is a line of future research. In order to solve this issues, maybe the construction of model based in PDE is more useful. There are two direct ways to extend the results of this work into a PDE model: the first one is to try to relate the macroscopic variables of the mean field description with some new variables that can be measured in experiments, doing further approximations for variables such as $\left\langle p^{act}\right\rangle$ that are difficult to relate to experiments. This approach will lead to a simplified form of equation (2.39). The second option is to build a new model, using PDEs, to describe not only the density of bacteria but the concentration of AIs and PGs. In this line, we could use the insight of the system we have gained via the individual based model to write a system of equations with parameters that can be related to experiments. This work is a first step to construct a more tractable description of the system, in order to get a quantitative description of the effect of quorum sensing regulation of public good proudction on *Pseudomonas aeruginosa* growth.

Bibliography

[1] C. M. Waters and B. L. Bassler, "Quorum sensing: Cell-to-cell communication in bacteria," Annual Review of Cell and Developmental Biology, **21**, 319 (2005).

[2] S. A. West, S. P. Diggle, A. Buckling, A. Gardner, and A. S. Griffin, "The social lives of microbes," Annual Review of Ecology, Evolution, and Systematics, **38**, 53 (2007).

[3] S. P. Diggle, A. S. Griffin, G. S. Campbell, and S. A. West, "Cooperation and conflict in quorum-sensing bacterial populations," Nature, **450**, 411 (2007).

[4] B. Allen, J. Gore, and M. A. Nowak, "Spatial dilemmas of diffusible public goods," eLife, **2**, e01169 (2013).

[5] W. W. Driscoll and J. W. Pepper, "Theory for the evolution of diffusible external goods," Evolution, **64**, 2682 (2010).

[6] B. Momeni, A. J. Waite, and W. Shou, "Spatial self-organization favors heterotypic cooperation over cheating," eLife, **2**, e00960 (2013).

[7] S. Heilmann, S. Krishna, and B. Kerr, "Why do bacteria regulate public goods by quorum sensing? how the shapes of cost and benefit functions determine the form of optimal regulation," Frontiers in Microbiology, **6**, 767 (2015).

[8] J. D. Dockery and J. P. Keener, "A mathematical model for quorum sensing in pseudomonas aeruginosa," Bulletin of Mathematical Biology, **63**, 95 (2001).

[9] J. P. Ward, J. R. King, A. J. Koerber, P. Williams, J. M. Croft, and R. E. Sockett, "Mathematical modelling of quorum sensing in bacteria." IMA Journal of Mathematics Applied in Medicine and Biology, **18**, 263 (2001).

[10] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman and Hall, 2007).

[11] A. Pai, Y. Tanouchi, and L. You, "Optimality and robustness in quorum sensing (qs)-mediated regulation of a costly public good enzyme," Proceedings of the National Academy of Sciences, **109**, 19810 (2012).

[12] A. K. Bhardwaj, K. Vinothkumar, and N. Rajpara, "Bacterial quorum sensing inhibitors: Attractive alternatives for control of infectious pathogens showing multiple drug resistance," Recent Patents on Anti-Infective Drug Discovery, **8**, 68 (2013).

[13] J. P. Gerdt and H. E. Blackwell, "Competition studies confirm two major barriers that can preclude the spread of resistance to quorum-sensing inhibitors in bacteria," ACS Chem. Biol., **9**, 2291 (2014).

[14] R. Toral and P. Colet, *Stochastic Numerical Methods: and introduction for students and scientists* (Wiley-VCH, 2014).

[15] C. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer-Verlag, 1985), 2nd edition.

[16] N. V. Kampen, *Stochastic processes in Physics and Chemistry* (Elsevier, 2007).

[17] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," Journal of Computational Physics, **22**, 403 (1976).

[18] D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," J. Phys. Chem., **81**, 2340 (1977).

[19] E. Heinsalu, E. Hernández-García, and C. López, "Competitive brownian and lévy walkers," Phys. Rev. E, **85**, 041105 (2012).

[20] M. Abel, A. Celani, D. Vergni, and A. Vulpiani, "Front propagation in laminar flows," Phys. Rev. E, **64**, 046307 (2001).

[21] Cecconi, F., Vergni, D., and Vulpiani, A., "Reaction Spreading in Systems With Anomalous Diffusion," Math. Model. Nat. Phenom., **11**, 107 (2016).

[22] R. Martínez-García, C. Murgui, E. Hernández-García, and C. López, "Pattern formation in populations with density-dependent movement and two interaction scales," PLOS ONE, **10**, e0132261 (2015).

[23] M. A. Nowak, "Five rules for the evolution of cooperation," Science, **314**, 1560 (2006).

[24] W. R. Young, A. J. Roberts, and G. Stuhne, "Reproductive pair correlations and the clustering of organisms," Nature, **412**, 328 (2001).

[25] A. Mund, S. P. Diggle, and F. Harrison, "The fitness of pseudomonas aeruginosa quorum sensing signal cheats is influenced by the diffusivity of the environment," mBio, **8**, e00353 (2017).

A. The Hole Model

In Section 1.3.2, we discussed about reaction-diffusion equations. During this discussion, we said that the stochastic reaction

$$A + R \xrightarrow{k} 2A, \tag{A.1}$$

is equivalent to the reactions (1.1) and (1.2) at the mean field level (even when it has differences in the fluctuations). So the natural question is to wonder why our model has to be ruled by (2.1) and (2.2) and not by (A.1). From a population dynamics point of view, reaction(A.1) means that if A reproduces if it has enough resources and space R, which we call «holes». In order to have something equivalent to (1.1) and (1.2), we need that every individual has its own rate k_j. Also, we need a variable number of holes, since the carrying capacity of (1.1) and (1.2) is a variable that depends on the mean growth rate. Then, we propose that the number of holes is $r = r(k) = N(k) - n$, where $N(k)$ is the carrying capacity. In the case of our model, we have $N(k) = k/\delta$, so in this case we use the ansatz $N(k) = \lambda k$ in an analogous way. The stochastic reaction equivalent to our model is given by,

$$A_j + R(k) \xrightarrow{k_j} A_j + A_{j+1}. \tag{A.2}$$

We only have one global growth rate, which is the product of $r(k)$ by every k_j,

$$\Omega(n \to n+1) = \frac{(N(k) - n)}{V} \sum_{j=1}^{n} k_j. \tag{A.3}$$

We have to compute $\langle \Omega(n \to n+1) \rangle$, which will lead directly to the mean field equation.

$$\frac{d\langle n \rangle}{dt} = \left\langle \frac{r(k)}{V} \sum_j k_j \right\rangle = \frac{1}{V} \langle n r(k) k \rangle \simeq \frac{1}{V} \langle n \rangle \langle r(k) \rangle \langle k \rangle . \tag{A.4}$$

Remember that $r(k) = \lambda k - n$, so $\langle r(k) \rangle = \lambda \langle k \rangle - \langle n \rangle$. To compute the average values over k, we get the expression (2.17), and apply the relation (2.6) for the amount of PG detected. Then we follow the same procedure described in Section 1.2.2 to compute the averages of the quantities. We assume also that the factor $1/V$ can be absorbed with a change of scale. Following this procedure, and assuming that all bacteria of the system are active, we get the piecewise defined equations,

$$\begin{cases} \frac{d\langle n \rangle}{dt} = \lambda k_{eff}^2 \langle n \rangle \left(1 + \frac{2\alpha P \lambda - 1}{\lambda k_{eff}} \langle n \rangle + \frac{\alpha P(\alpha P \lambda - 1)}{\lambda k_{eff}^2} \langle n \rangle^2 \right), & n^{act} < u_{sat} \\ \frac{d\langle n \rangle}{dt} = \lambda \left(k_{eff} + \alpha u \right)^2 \langle n \rangle \left(1 - \frac{\langle n \rangle}{\lambda (k_{eff} + \alpha u)} \right), & n^{act} \geq u_{sat} \end{cases} \tag{A.5}$$

where $k_{eff} \equiv k_0 - \beta q$ to make the notation lighter. See that when $n^{act} \geq u_{sat}$ we have a logistic like equation with a carrying capacity that is exactly the same as in our original model, letting $\lambda = 1/\delta$. However, the growth rate of this model is much faster, since there is no death in the system. In addition to that, we can

see that in the divergent phase a term $\langle n \rangle^2$ appears. This is important because the divergence in our model is different from the one in this hole model. In the original one, we had that it grew as $\langle n \rangle^2$, but in this case the divergence is of type $\langle n \rangle^3$. Moreover, $\alpha P/\delta > 1$ in practically all cases, meaning that this will be always positive, giving a fast growth up to the saturation.

We include this model here because from the analysis of the differences is interesting, since they arise when we introduce k_j and the variable carrying capacity $N(k)$, but it is exactly the same as the logistic for constants k and N. We considered this model also for the growth of bacteria, but there are two points that made us discard it:

1. The growth has some slight differences with a pure logistic, that describes very well the growth of bacteria. The growth rate of this model is too fast, specially in the divergent stage.

2. In this model, bacteria cannot die, so it is impossible to get results as the ones in 2.5 for the rich medium, where they have a decrease in the number of bacteria. The number of individuals can only increase, so it makes it inappropriate from the biological point of view.

B. Mean field for simple competition

We derived the mean field equation for the reactions (1.1) and (1.2), that turned out to be a logistic equation. Then we introduce modifications to this reactions and analysed the problem for single species in our model. In this Appendix we want show dynamics of the competition of mixed populations for the simple model with constant rates and carrying capacity, that will give us some insight into our model. Let us think we have two different kind of bacteria. Then, (1.1) and (1.2) can be extended to,

$$A \xrightarrow{k_1} 2A, \qquad B \xrightarrow{k_2} 2B, \tag{B.1}$$

$$A + A \xrightarrow{\delta} A, \qquad B + B \xrightarrow{\delta} B, \tag{B.2}$$

$$A + B \xrightarrow{\delta} A, \qquad A + B \xrightarrow{\delta} B. \tag{B.3}$$

Note that A and B grow with different rate k_1 and k_2, and all the deaths happen with the same rate δ. If we call a and b the number of individuals of each kind, then the global rates for the system are given by,

$$\Omega\left[(a,b) \to (a+1,b)\right] = k_1 a \tag{B.4}$$

$$\Omega\left[(a,b) \to (a,b+1)\right] = k_2 b \tag{B.5}$$

$$\Omega\left[(a,b) \to (a-1,b)\right] \simeq \frac{\delta a}{V}(a+b) \tag{B.6}$$

$$\Omega\left[(a,b) \to (a,b-1)\right] \simeq \frac{\delta b}{V}(a+b) \tag{B.7}$$

Note that for the death processes we have to sum the contribution of all the reactions that involve losing one individual of that type. We redefine $\delta \equiv \delta/V$. Equation (1.9) can be readily generalized for several variables as

$$\frac{\partial p\left(\vec{n},t\right)}{\partial t} = \sum_{\left\{\vec{\ell}\right\}} \left(E^{\vec{\ell}} - 1\right)\left[\Omega\left(\vec{n} \to \vec{n} - \vec{\ell}\right) p\left(\vec{n},t\right)\right], \tag{B.8}$$

where $\left\{\vec{\ell}\right\}$ is the set of the vectors we use for our global rates, and $E^{\vec{\ell}}\left[f\left(\vec{n}\right)\right] = f\left(\vec{n} + \vec{\ell}\right)$. For example, in our case, $\left\{\vec{\ell}\right\} = \left\{(-1,0),(0,-1),(1,0),(0,1)\right\}$. Using this generalization, we can compute the system of equations for the first moment,

$$\frac{d\left\langle\vec{n}\right\rangle}{dt} = -\sum_{\left\{\vec{\ell}\right\}} \left\langle \vec{\ell} \cdot \Omega\left(\vec{n} \to \vec{n} - \vec{\ell}\right)\right\rangle. \tag{B.9}$$

Note that the global rate is a scalar quantity. That means that for the equation of $\langle a\rangle$ will only contribute terms with $\ell_x \neq 0$. In fact, there are two terms for the equation of $\langle a\rangle$ and another two for the equation of $\langle b\rangle$, as we could expect,

$$\begin{cases} \frac{d\langle a\rangle}{dt} = k_1\langle a\rangle - \delta\left\langle a\left(a+b\right)\right\rangle, \\ \frac{d\langle b\rangle}{dt} = k_2\langle b\rangle - \delta\left\langle b\left(a+b\right)\right\rangle. \end{cases} \tag{B.10}$$

Now, we have to apply the mean field approximation. In addition to the elimination of the fluctuations, $\langle a^2 \rangle = \langle a \rangle^2$, we will suppose that the values of variables a and b are not correlated. This is only an approximation, and if we decided to go for higher orders, we would need equations for $\langle a^2 \rangle$, $\langle b^2 \rangle$ and also for $\langle ab \rangle$. However, if we use the mean field approximation the system is closed,

$$\begin{cases} \frac{d\langle a \rangle}{dt} = k_1 \langle a \rangle \left(1 - \frac{\delta}{k_1} (\langle a \rangle + \langle b \rangle)\right), \\ \frac{d\langle b \rangle}{dt} = k_2 \langle b \rangle \left(1 - \frac{\delta}{k_2} (\langle a \rangle + \langle b \rangle)\right). \end{cases} \tag{B.11}$$

Equation (B.11) is again a logistic equation with a different carrying capacity $N_j = k_j/\delta$ for each population of the system. If we compute the fixed points of the system, we get

$$\begin{aligned} \vec{a}^* &= (N_1, 0), \quad \text{(stable if } k_2 < k_1\text{)}, \\ \vec{b}^* &= (0, N_2), \quad \text{(stable if } k_2 > k_1\text{)}, \end{aligned}$$

that means that the species with the higher growth rate will prevail in the system, and no coexistence is allowed. When the system has reached the carrying capacity for the strain that grow faster, say A, particles of kind B can die and then be substituted by particles of kind A. Death happens at the same rate for each individual, but it is more probable to have a birth of A than a birth of B, leading to $b = 0$ at large times if $k_1 > k_2$. The case $k_1 = k_2$ is marginally stable.

This simple example is extensible to our mixed populations complete model. We can easily extend equation (B.11) using similar arguments as the ones we employed in Section 1.2.2. Basically this equation allows us to understand the basic dynamics of the mixed populations: the strain with higher growth rate is going to reproduce faster. When the total carrying capacity of the system is reached, this strain still can grow up to a equilibrium point that is determined only by the relations of the growth rates. In this case it is more complicated since the growth rates can vary, as well as the carrying capacities.