Stochastic Models of Microbial Communities: Stochastic Dynamics of Quasi-Neutral Species in a Resource-Limited Chemostat Environment

Siding Luo

School of Engineering

University of Glasgow

A thesis submitted in fulfilment of the requirements

for the degree of

Doctor of Philosophy

To my parents

谨此献给我的父母

Acknowledgements

First of all, I would like to express the great appreciation to my supervisor, Dr. Christopher Quince, who offered me this precious opportunity to do research on the stochastic dynamics in microbial communities four years ago. The way of scientific researching is full of challenges, but as my research advisor, his positive attitude, valuable advice and tremendous patience helped me survive the toughest period in my life. His high standard significantly improves my research ability and work over this whole research course. I am deeply grateful to him for going through this thesis again and again, correcting my writings and offering numerous suggestions.

I would also would like to give my appreciations to Dr. Todd Parsons for his great preliminary work in the stochastic population dynamics, helpful ideas and constructive suggestions. Thanks for providing me plenty of support in understanding mathematical population dynamics and giving me inspiration to complete many analytical calculations and derivations, especially in deriving the diffusion approximation at long time scale and working out the solution to correct some unpredictable analytical results. He also helped check all the analytical calculations in Chapter 3 of this thesis.

I am grateful to Dr. Zofia Jones with her numerous discussions and suggestions. My thanks also to Dr Umer Ijaz for his expertise in programming, which helped me to sort out practical problems in simulation.

I wish to thank Prof. William Sloan for his kind encouragement, many kinds of help. I would also like to thank all the girls in my office for sharing the pleasant environment.

I am immensely grateful to the Department of Civil Engineering, the University of Glasgow, and Engineering and Physical Sciences Research Council (EPSRC) for awarding me the Scholarship.

Last but definitely not least, particularly thankful to my parents and husband, without your consistent support, encouragement and love, I would never know when I could complete this thesis work.

Abstract

The most indispensable work for microbial ecologists is to develop mathematical models in order to describe microbial communities. In this aspect, a proper understanding of microorganism richness and abundance is of paramount importance. A chemostat environment is a classic open microbial community, where multiple species compete for limited nutrients, whose mathematical model has broad applications in microbiology and population biology.

Generally speaking, there are four key processes that may influence the diversity: selection, speciation, drift and dispersal. The debate between niche assembly theory and neutral theory has lasted for decades about the dominant process. In term of simplicity, Hubbell's unified neutral theory of biodiversity has a distinct advantage for sampling and parameterisation. It offers a quantitative stochastic base model of island macroscopic community coupled with meta-community, where species compete in a finite environment.

In this thesis, an explicit quasi-neutral chemostat model is fully devised by reconciling neutral theory and niche difference, which gives insight into

how the origin, maintenance and loss of biodiversity in the local competitive community at different time scales are influenced by selection, stochastic drift and dispersal. An analysis of the deterministic dynamics is conducted to explain the niche assembly rule, and to further show that the species at the same largest fitness will be selected to coexist through life history trade-offs. These species are quasi-neutral. However, over a long period of time, stochastic drift will play a dominant role in constructing the pattern of the local communities. Without dispersal, extinction is the ultimate fate of stochastic drift.

Both analytical and numerical methods established in this thesis verify that the quasi-neutral species are in fact not competitively equivalent. Their difference will drive a superior species to fix in the isolated community. When dispersal is incorporated, even with low immigration rate, it will drive the long term drift of large communities, and balance extinction to maintain the diversity of the local communities. An explicit results for the stationary abundance distribution is calculated, which helps to demonstrate a deviation from the neutral model. These results from the explicit stochastic model highlight the importance of incorporating species interaction into the basic neutral model.

Contents

List of Figures

List of Tables

Main Notation List

$\Rightarrow$ defines convergence in distribution.

$N(t)$ is the number of cells at time t.

Volume V is in the units of volumes of typical cell in analytical models.

Volume V is in the units of ml in simulation models.

$S(t)$ is the substrate concentration or resource concentration at time t (mass/volume).

$x(t)$ is the population density at time t (number/volume).

$\left(\hat{S}(t), \hat{x}(t)\right)$ are stochastic processes at time t.

$\left(\bar{S}(t), \bar{x}(t)\right)$ are deterministic processes at time t.

Ω=(S^*, x^*) is the equilibrium state in the deterministic model.

$\beta_l(t)$ is the jump intensity of change l at time t in birth and death process.

S^{in} is the input substrate concentration (mass/volume)

D is the dilution rate, (time^{-1})

m_k is the maximal growth rate of the kth species (time^{-1})

y_k is the yield coefficient of the kth species (cells/mass)

a_k is the half-saturation constant of the kth species (mass/volume)

μ_k is the death rate of the kth species (time^{-1})

$b_k(S,x) = \frac{m_k S}{a_k+S}$ is the birth rate of the kth species at state $(S(t), x(t))$.

$d_k = D + \mu_k$ is the death rate of the kth species.

$\gamma_k = \frac{\partial b_k(S)}{\partial S}|_{S^*}$ is the sensitivity of the birth rate to the resource at the resource concentration at the equilibrium state

R_k is the immigration rate (number/unit time)

Chapter 1

Introduction and Literature Review

1.1 Introduction

Population biology is the branch of mathematical biology that studies how birth, death, immigration and emigration change the size and age composition of populations over time, and how these changes are influenced by biological and environmental processes. There are essentially two streams of problems that are represented at the population level. One is population ecology, which is concerned with changes in population structure as a consequence of interactions of organisms with the environment, with individuals of their own species, and with those of other species. The other is population and evolutionary genetics, whose focus is on the frequency distribution and interaction of alleles and genes in populations.

Our research on microbial population biology applies population ecology

and population genetics in order to understand the evolution and ecology of microbes and the community interactions between microorganisms, including nutrient competition. In this thesis, a specific chemostat environment will be hypothesised in order to explore microbial ecology and evolution. Before we present the detailed mathematical model, this chapter provides a relevant review of the development of mathematical descriptions of population dynamics, and a summary of the historical debates between deterministic and stochastic views, as well as a clarification of the motivations and objectives of this thesis.

1.2 Review of the Population Dynamics

In the history of mathematical population dynamic models, nearly all concepts in population ecology and population genetics originated from the same processes and models, and diverged later due to their different application areas. But the similarity in modelling analogies may still be found in later development. The development of the concept of population ecology predates the understanding of genetics. Therefore, with the purpose of understanding the microbial community, a full retrospective view of the whole area of population dynamics is helpful. In this section, we will briefly discuss some history of population dynamics modelling relevant to this thesis, and present the long-lasting debates between deterministic and stochastic aspects.

1.2.1 Development of the Mathematical Population Dynamics

Fibonacci's rabbit problem proposed in 1202 is probably the first effort in the history of population dynamics. The Fibonacci sequence was created to model a population of rabbits, but was far from being realistic, since no mortality and no separation of sexes, etc. were assumed. Half a millennium later, Euler introduced the geometric (exponential) growth of populations [14], where the population P_n at year n increases geometrically,

$$P_n = (1 + x)^n P_0,$$

with year n, initial population P_0, and growth rate x. Euler's calculations showed that a population could increase more than tenfold within one century, which precisely matched with observations at the time in the city of London [5].

However, with Euler's geometric growth equation, even a small initial population growing by a tiny constant fraction will lead to a large unsustainable population on Earth. Half a century later, this question was addressed by Thomas Robert Malthus. He published one of the earliest and most influential books on populations, *An Essay on the Principle of Population* [41]. The views in the book suggested that growing population rates would contribute to a rising supply of labour that would form a kind of resistance (inevitably lower wages). His work was a key influence on Charles Darwin and Alfred Russell Wallace's development of the theory of natural selection [11].

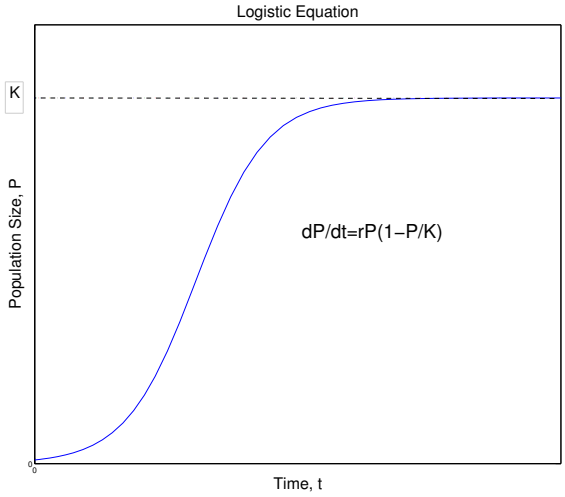

Figure 1.1: Population growth under the Verhulst logistic equation is sigmoidal (S-shaped), reaching an upper limit termed the carrying capacity, K. Populations initiated at densities above K decline exponentially until they reach K, which represents the only stable equilibrium.

Adolphe Quetelet and Pierre-Francois Verhulst continued this problem. They translated the phenomena that a population approaches a steady state into a mathematical model with the population expressed by $P(t)$ at time t,

$$\text{Logistic equation:} \qquad \frac{dP}{dt} = rP\Big(1 - \frac{P}{K}\Big), \tag{1.1}$$

where the constant r defines the growth rate and K is the carrying capacity.[1] Due to the S-shape as illustrated in Figure 1.1, the name "logistic equation" was given in [68]. The idea of density dependence was first introduced by Verhulst. In his model, the logistic equation could be approximated by

[1] The carrying capacity of a biological species in an environment is defined as the maximum population size of the species, which is different from the concept of population equilibrium.

the exponential equation, while P is less than K in the initial phase. As saturation begins (due to environmental pressures), growth slows down. At maturity, growth stops and the population is maintained at the stationary state. The "S shaped" population curve is obtained. This is one of the most basic models of population growth and today is considered a milestone in this history [5].

In 1875, the Galton-Watson process was presented by Francis Galton and Henry William Watson while analysing the problem of extinction of family names [72]. This model accurately describes Y chromosome transmission in genetics. The Galton-Watson process is a branching process $\{X_0, X_1, \ldots, X_N\}$, where X_k is the number of male members in the k-th generation. By assumption, $X_0 = 1$, and $\xi_j^{(k)}$ is the number of male descendants of the j-th member in the k-th generation, with probability function $P(\xi_j^{(k)} = i) = p_i$. Therefore, $X_{k+1} = \sum_{j=1}^{X_k} \xi_j^{(k)}$ and $\sum p_i = 1$. To calculate the probability of extinction of the family name within n generation, a trick was used by Waston. He considered the generating function for the first generation $f_1(x) = p_0 + p_1 x + p_2 x^2 + \cdots + p_k x^k$, and noticed that the n^{th} generation $f_n(x)$ could be computed recursively from the formula

$$\begin{aligned} f_n(x) &= p_{0,n} + p_{1,n}x + p_{2,n}x^2 + \cdots + p_{q^n,n}x^{q^n} \\ &= f_{n-1}\Big(f_1(x)\Big) \end{aligned}$$

where $p_{k,n}$ is the probability that generation n consists of k man [72]. As a result, the probability of extinction of the family name within n generations $p_{0,n} = f_n(0)$ is yielded.

From 1900, the rediscovery of the laws of Mendelian inheritance initiated the modern science of genetics. The Hardy-Weinberg principle (1908) stated that in the absence of selection, mutation, migration and genetic drift, the allele frequencies in a large population were constant through generations [27]. Taking for instance two alleles, A and a, dominant and recessive respectively, Hardy determined the frequencies of the genotypes AA, Aa and aa, staying constant through the generations, as equal to p, $2q$ and r respectively, satisfying the condition $q^2 = pr$. This law for gene frequencies was also discovered by Weinberg in the same year [76].

After more than six centuries of development, mathematical biologists needed more complex models to describe the interactions between species. The Lotka-Volterra equations were then proposed by Alfred James Lotka [39] and Vito Volterra [70], independently, to describe the dynamics of ecological systems with predator-prey interactions, competition, diseases and mutualism [5]. In the Lotka-Volterra equations, the growth rates of two populations prey $N_1(t)$ and predator $N_2(t)$, at time t can be expressed by the equations [39],

$$\begin{aligned} \frac{dN_1}{dt} &= N_1(\alpha_1 - \beta_1 N_2) \\ \frac{dN_2}{dt} &= N_2(-\alpha_2 + \beta_2 N_1), \end{aligned}$$

where the parameter α_1 is the intrinsic growth rate of prey, α_2 is the intrinsic death rate of predator, and $\beta_1 N_1 N_2$ and $\beta_2 N_1 N_2$ represent the rate of predation upon the prey and growth of the predator population respectively, both proportional to the rate at which the predators and the prey meet. By

studying the sign of dN_k/dt, Lotka explained the four equilibrium states, why and also how the system could oscillate in a periodic way. This model is also widely applied to determine the population dynamics in the case of an epidemic, where $\beta S(t)I(t)$ could express the rate of diseases spreading from infected population $I(t)$ to susceptible population $S(t)$ at time t, with infection parameter β.

The discipline of population genetics was founded in 1930 through the work of R.A. Fisher, J.B.S. Haldane and Sewall Wright. Fisher made a key step in population genetics by reconciling Mendel's law with the ideas of natural selection emphasised by Darwin [5]. Fisher explained the two situations of coexistence or extinction of genotypes, and published the book *The Genetical Theory of Natural Selection* [20] in 1930, which was a milestone in the history of population genetics. Influenced by Fisher, Wright considered a mathematical model following the Hardy-Weinberg principle, but without assuming the population to be infinitely large [77][78]. He first introduced the concept of "drift",[2] which questioned natural selection as the foundation of evolutionary theory, and ignited a century-long debate over the respective influence of natural selection and random drift.

Most multicellular organisms are diploid, that is they have two sets of chromosomes and have one copy of each gene (one allele) on each chromosome. If there is N copies of genes, the whole population of the allele is $2N$. Assume A and a are the two alleles segregating in the population. Under the assumption of random mating, the Wright-Fisher Markov process calculates

[2] The concept of "drift" in ecology is the change in the frequency of an allele/individual in a population due to random sampling.

the transition probability $P_{i,j}$ of obtaining j copies of the allele A that had i copies in the last generation, which is

$$\text{Wright-Fisher process:} \qquad P_{i,j} = \binom{2N}{j}\Big(\frac{i}{2N}\Big)^{j}\Big(1-\frac{i}{2N}\Big)^{2N-j}.$$

Using the binomial distribution, for each generation, the number of copies of A is expected to remain the same on average, but in fact may take any value from zero (become extinct) to $2N$ (fixation) [58][73].

While the Wright-Fisher model has sudden-death generations, Moran (1958) [46] embedded the Markov process in continuous time by assuming overlapping generations.[3] In this case, when an individual dies, it is immediately replaced by a newcomer. Assume there are i copies of allele A and $2N-i$ copies of allele a in the current generation, and let j be the number of copies of allele A after one time unit. The transition probabilities $P_{i,j}$ are,

$$P_{i,j} = \begin{cases} \frac{i}{2N}\frac{2N-i}{2N} & \text{if } j = i+1 \\ \frac{2N-i}{2N}\frac{i}{2N} & \text{if } j = i-1 \\ (\frac{i}{2N})^2 + (\frac{2N-i}{2N})^2 & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases}$$

Supposing the population $2N$ to be large, the mean and the variance of the copies of allele A, $X(t)$ at time t, can be detained when an initial state

[3] The non-overlap in generations means each generation produces exactly one generation in each time period.

$X(0) = i$ is given:

$$E[X(t)|X(0) = i] = i$$
$$Var[X(t)|X(0) = i] = 2\frac{i}{2N}\frac{2N-i}{2N}\Big(\frac{1-(1-\frac{2}{(2N)^2})^t}{\frac{2}{(2N)^2}}\Big).$$

Moran's model allows us to derive many exact results that are available only approximately under the Wright-Fisher model. In practice, with equivalent definitions of a generation, the Moran model and Wright-Fisher model give qualitatively similar results, but genetic drift runs twice as fast in the Moran model as it does in the Wright-Fisher model [73].

Two main findings emerges from the Wright-Fisher model and Moran's model. First, if the population is assumed infinitely large, the fluctuations of the allele result in time-varying genetic drift without changing the main qualitative features of the conclusions. Second, if one takes the finite size of the populations into account, then genetic drift plays a significant role. The frequencies of the two alleles fluctuate and one of the alleles will be lost from the population. An excellent treatment of the dynamics of this model are given by Ewens in [16]. In our later analysis in this thesis, with the help of the law of large numbers, the first conclusion (that a small demographic drift in a large population will not change the deterministic limit) is easily proved. Using the simulation results in small population size, the second conclusion (that a significant role is played by randomness) is also proved.

The Wright-Fisher model shows its robustness in the later development of population genetics, and most mathematical descriptions of allele frequencies are built upon it. The coalescent process, which was originally

developed in the early 1908s by John Kingman is a backward view of the time evolution of the Wright-Fisher model when the population size is large, and describes the genealogy of a population sample [73]. Given a Wright-Fisher model with population size $2N$, the ancestral process includes both a discrete genealogical tree structure and coalescent time intervals. Due to the random picking of a parent, the probability that two individuals choose the same parent is $p_{c,1} = \frac{1}{2N}$. The probability for coalescence of two individuals i.e. a common ancestor for the first time at t generations ago, is then

$$p_{c,t} = \frac{1}{2N} \cdot (1 - \frac{1}{2N})^{t-1}.$$

Therefore, the coalescent time t is geometrically distributed with success probability $\frac{1}{2N}$ [58].

Kimura (1962) later extended the results of the Wright-Fisher model to include the probability of eventual dominance, and presented a general diffusion approximation formula of the fixation probability $u(p,t)$ in terms of the initial frequency p,[4]

$$\frac{\partial u(p,t)}{\partial t} = \frac{V}{2}\frac{\partial^2 u(p,t)}{\partial p^2} + M\frac{\partial u(p,t)}{\partial p}$$

where V and M are the variance and mean of the change of p per generation [35]. This partial differential equation (PDE) is known as the backward Kolmogorov equation and can be solved with boundary conditions $u(0,t) = 0, u(1,t) = 1$ to obtain the probability $u(p,t)$. Kimura's standard

[4]In population genetics, fixation is the change in a gene pool from a situation where there exists at least two variants of a particular gene (allele) to a situation where only one of the alleles remains.

diffusion approximation has made an enormous impact on the development of theoretical and applied population genetics and was made more robust by Ewans [52][16].

All the population genetics models built upon the Wright-Fisher model assume a fixed population size, where the population $2N$ is an even number and has no mathematical significance [46]. However, in a real microbial community, a fixed population is an unrealistic assumption. By varying stochastically the population size, Quince and Parsons developed their density dependent model for the population genetics [52][53][54]. In their work, the birth and death rates of competitive species are defined according to the total population size as follows:

$$\begin{aligned} Pr(X_i + 1|X_i) &= \beta_i X_i \\ Pr(X_i - 1|X_i) &= \delta_i \Big(1 + \frac{\sum_{j=1}^{K} X_j}{N}\Big) X_i, \end{aligned}$$

where β_i, δ_i are the intrinsic birth and death rate of species type i, and N is a parameter denoting the typical population size. Afterwards, a new diffusion approximation is introduced to admit qualitatively different diffusion from the Kimura's standard one, which demonstrates novel behaviour by contrasting small populations with populations near equilibrium. In order to derive this diffusion approximation, Parsons further developed a method by applying Itô's formula on a different time scale [51].

1.2.2 Debates between Deterministic and Stochastic Views in Community Ecology

Opinions in both population genetics and ecology on the role of stochastic factors are divided. In population genetics, a long-standing debate has persisted over whether most changes in gene frequencies result from randomness, neutral evolution or from natural selection. Since this thesis is on ecology in microbiology, we will focus in this subsection on reviewing the debates in community ecology.

There are two conflicting schools in ecology in terms of the importance of the role that drift and dispersal play in ecological communities. The mainstream perspective is that of traditional niche-based theories which inherit the Darwinian tradition, believing that species differ in their traits, focusing on the lives of interacting individuals and their fitness consequences [11]. In this school of thought, biodiversity can be determined by the "niche assembly rules", which are based on the ecological niches or functional roles of each species. Coexistence in this niche-assembly view is an equilibrium state with the best competitors occupying each ecological niche, and this perspective predicts the balance among these niche-differentiated competitor species, predators and their prey [67].

In contrast to deterministic coexistence and diversity owing to competitive interactions, some theorists have questioned whether similar patterns could be observed through random chance alone. The other perspective is formed by believing that niche differences quickly screen out the inferior species and these differences are never the leading mechanism in controlling

the abundance of species over a long period of time. Instead, ecological drift, which emphasises the role of stochastic processes, plays a prominent role in shaping the diversity of the community. The coexistence of species results from a balance between speciation/ immigration and extinction through random demographic drift. The emphasis on the ecological drift leads to a theory with species that are ecologically equivalent [60] [59] [67]. MacArthur and Wilson's island biogeography theory in Figure 1.2 is a model of this type.[5] The actual parameters of their model are immigration and extinction rates, distance from mainland source areas and island size, parameters which are absent from most niche-assembly theories [31]. However the most extraordinary work was developed by Stephen Hubbell, who proposed that the abundances and diversity of species in a community are determined mainly by random dispersal, speciation, and extinction, with the assumption of neutrality on the individual level which means the 'whole population one niche'.

This controversy between niche assembly deterministic theory and stochastic neutral theory has lasted for half a century. Although niche difference is a mechanism that can maintain biodiversity by allowing species to coexist, the classical niche assembly theory does not provide a general explanation for the relative abundance of species [31][60], and only a limited number of analytical models in niche assembly theory can be applied to a small sample [47][65]. Stochastic drift should be taken into account in any attempt to understand population structure, but the assumption of ecologically equivalent

[5] An island in this context, is not just a segment of land surrounded by water. It is any area of habitat surrounded by areas unsuitable for the species on the island.

species is obviously the great weakness of neutral theory. As a consequence, more and more synthetic theories are and will be developed to reconcile ecological drift and niche differences among species.

The chemostat model we built in this thesis is a such a reconciliation, where niche differences, stochastic drift and dispersal interact to shape the local microscopic community. For that reason, firstly we will introduce the neutral theories in macroscopic communities in Section 1.3.

1.3 Neutral Theory in Community Ecology

In community ecology, 'neutral' means that a theory treats organisms as identical in their probabilities of giving birth, dying, migrating, and speciation. Each theory uses a definition of neutrality operating on a different level. Although controversial, these neutral theories demonstrate how much can be achieved with the simple assumption of ecological equivalence before introducing more complexity.

1.3.1 Neutral Models in Macroscopic Community Ecology

To better understand the stochastic dynamics of biodiversity in community ecology, Caswell (1976) attempted to emphasise the role of drift by incorporating the neutral approach from population genetics without considering the complex interactions and environmental effects. Unfortunately he failed to provide a good fit to the observed data [7]. It was MacArthur and Wilson who first erected a new radical theory, separated from the mainstream Darwinian view of ecological communities, by assuming neutrality

and dispersal-assembly, emphasising the importance of random dispersal and ecological drift in the communities.

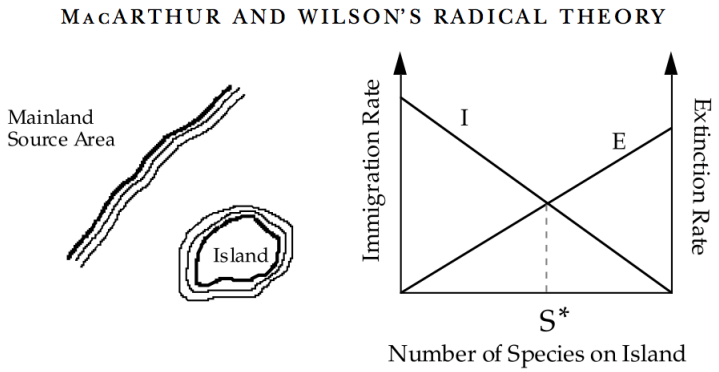

Figure 1.2: MacArthur and Wilson's island biogeography theory explains the number of species on islands as a dynamic equilibrium (S^*) between the rate of immigration of new species onto the island and the rate of extinction of species already resident on the island. I is immigration rate and E represents extinction rate.

In MacArthur and Wilson's island biogeography model in Figure 1.2, neutrality is defined at the species level, and immigration and extinction rates are introduced as parameters which are absent from niche-assembly theories. Although a steady-state number of species on the island is maintained through repeated immigrations and local extinctions, a stable assemblage of particular taxa cannot be predicted. This means that the equilibrium state only exists in a narrow sense, and the island biogeography model is not a equilibrium theory. Thus, the absence of speciation and relative abundance of species makes it incomplete as an ideal dispersal assembly theory.

1.3.1.1 Hubbell's Unified Neutral Theory of Biodiversity

By integrating speciation into the island biogeography model and by moving the neutrality from species level to individual level, Hubbell constructs his neutral theory of biodiversity and biogeography, which assumes that ecological communities are structured entirely by ecological drift, random migration, and random speciation [31].

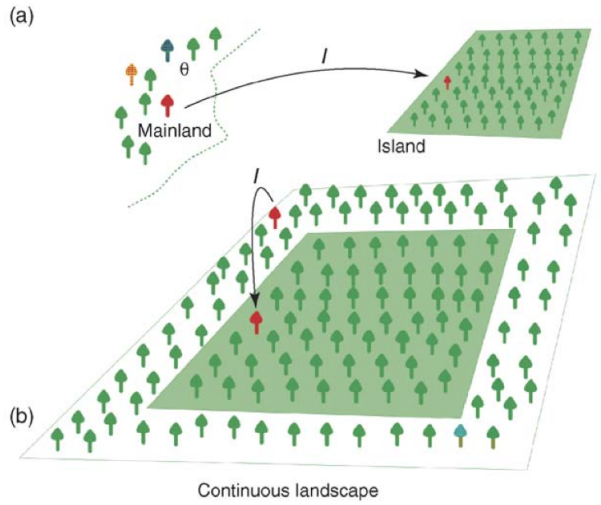

Figure 1.3: This diagram of Hubbell's mainland-island model and its relation to a sample from a continuous landscape is from [3]. In diagram (b), the green shaded rectangular area is the local community, within a continuous meta-community. The fundamental biodiversity and dispersal numbers in (b) can be approximated by calculating the effective neutral mainland-island model in (a). The fundamental biodiversity number θ is a measure of the effective meta-community diversity, while the fundamental dispersal number I is a measure of the effective degree of isolation of the local community.

In his model, Hubbell defines an ecological community as a group of tropically similar species that actually or potentially compete in a local area for the same or similar resources. Neutrality, defined at the individual level,

is used to describe the assumption of per capita ecological equivalence of all individuals of all species in the community. He makes the definition of "biodiversity" to be synonymous with species richness and relative abundance in space and time, and quantitatively predicts the species presence-absence, relative abundance and persistence times in a wide range of different biological communities.

There are two sample size independent fundamental parameters in the theory that effectively determine the steady-state distribution of the species' richness: the fundamental biodiversity number, which is a measure of the effective meta-community diversity, and the fundamental dispersal rate (immigration rate m), which is a measure of the effective degree of isolation of the local community.

Hubbell places great emphasis on the argument that living nature is forever locked in a saturated, life-or-death and zero-sum game for limited resources. This assumption of saturation keeps the population constant. In local communities, the vacancy created by the death of an individual will be filled immediately by random birth within the local community with probability $1 - m$, or an immigration from the meta-community with probability m. Given the neutrality at individual level, the probability that a new individual comes from species i is simply proportional to that species' overall abundance, and is independent of species identity. Consequently, the expressions for the effective birth and death rate of the i^{th} species in the local

community are

$$Pr(N_i + 1/N_i) = (1-m)\frac{N_i}{J}\frac{J-N_i}{J-1} + m\frac{\mu_i}{J_m}\Big(1-\frac{N_i}{J}\Big) \tag{1.2}$$

$$Pr(N_i - 1/N_i) = (1-m)\frac{J-N_i}{J}\frac{N_i}{J-1} + m\Big(1-\frac{\mu_i}{J_m}\Big)\frac{N_i}{J} \tag{1.3}$$

where J and J_m are the population sizes for the island and meta-community respectively, and N_i and μ_i are the abundance of the i^{th} species in the island and meta-community respectively [31]. In his famous book [31], Hubbell provides a numerical calculation method using a generator function to generate one random sample.

Following on from Hubbell's basic theory, McKane et al. provide a general analytic solution by recasting Hubbell's discrete stochastic theory as an appropriate continuous Markovian process, and determine that the stationary probability distribution in local communities can be expressed by a beta function [44]. Moreover, to understand how the community is assembled from a given starting point, studying the stationary density is not enough. Since it is impossible to explicitly calculate the distribution of species at time t, it is necessary to consider a good analytical diffusion approximation. Using Van Kampen's system size expansion [34], Mckane et al. approximate the probability distribution involved with time to a Gaussian expression [44].

Despite its radical assumptions, Hubbell's Unified Neutral Theory of Biodiversity (UNTB) is the first general quantitative theory which attempts to explain the distribution, abundance and diversity of species, and gives insight into the origin, maintenance and loss of biodiversity in a biogeographical context by making only a few, but fundamental, assumptions [3].

No other alternative model has done this. It is still the only sampling model which has been developed. Although large numbers of empirical tests of the prediction of the neutral theory appear to have rejected this pure neutrality [42][43][60], the really useful point is that it can be considered as a null hypothesis to compare the actual data and assess the influence of other parameters [66].

1.4 Motivations and Objectives of the Thesis: Demographic Competition Models in a Resource-Limited Chemostat Environment

In a microbial community, the disparity between sample scale and community size is huge. Bacterial richness far exceeds the richness levels observed in the macroscopic community. Hundreds of thousands of bacterial taxa exist in a tiny sample and increase with the sample size [19]. For example, as many as 10^9 individual microorganisms can be found in a gram sample of soil [63]. All these factors make the attempt of ecologists to capture the empirical definition of a taxa-abundance distribution in nearly any community, and to derive analytical predictive models capable of dealing with very large populations and communities, very challenging [19][63][75]. Therefore, even with the critical assumption of neutrality, Hubbell's neutral model is still considered to be an appropriate quantitative base model for parameterisation and prediction.

The objective of this thesis is to construct suitable models, which are

not only capable of capturing the essential biological features of microbial population in compact communities, but are also mathematically tractable.

In Hubbell's Unified Neutral Theory of Biodiversity, he gives an example illustrating how species are distributed widely and defining the field of macro-ecology [31], however, he does not take species interaction into account but simply assumes ecological neutrality on the individual level. Among various methods of introducing niche structure into Hubbell's neutral model, the simplest one is to alter this assumption and make the birth, death and dispersal rate of species i dependent on the identity of i [7]. Sloan frees this assumption of the uniform birth rate for each species with selective parameters to model species abundance distributions in a prokaryotic community by applying the forward Kolmogorov equation [62][63]. Under the limit of neutrality, the beta-distributed stationary probability density of the relative abundance is analytically proved and is also recovered through numerical calculations. Compared to other works on the analytical solution advancing Hubbell's original model [30][66][69], Sloan's method not only redefines Hubbell's strict neutrality to satisfy different modelling demands, but also gives an straightforward diffusion approximation in a large population transferred from the original discrete model. It helps us to predict the dynamics of microbial communities using all the methods from modelling stochastic macroscopic community and population genetics. However, it is still based on the neutral theory and thus lacks an explicit demographic ground in ecological theory. Without factoring in some important mechanisms, it may mislead the interpretation and result in incomplete conclusions about the fundamental drivers of population dynamics and species diversity

[2][21][38]. Further iterations of candidate models for the prediction of theoretical microbial ecology are hence necessary.

In this thesis, we will redefine the assumptions of zero-sum and neutrality, build a niche-based stochastic demographic model for the multiple species case in a resource-limited chemostat environment to assess the influence of the mechanisms of selection, drift and dispersal, analyse the diversity over different time scales, and compare the results with those of Hubbell's neutral model and Sloan's near-neutral model to determine whether the species with same fitness exhibit substantial difference.

Using the Monod growth function, Hubbell developed a deterministic model to analyse species competition analytically in a single nutrient-limited chemostat environment and predicted that the single species with strongest fitness will survive [33]. Later Hubbell confirmed the agreement between this theoretical prediction and experimental outcomes [26].

To build our model, the same growth function as in Hubbell's chemostat model [33] is used. While keeping the growth and removal events of nutrient happening deterministically, demographical stochasticity is incorporated into the growth and death events of the population. A hybrid model, in which the birth and death processes of species are time-dependent through a nutrient concentration that is continuously evolving and constrained by the population size, is then built. Life history trade-offs [6] allow different species to survive in Hubbell's selection regime.

Collet and Melerd [10] are concerned with a monotype population in

[6] Life history trade-offs equalise the per capita relative fitness of species in the community, which set the stage for ecological drift [31].

chemostat model and discuss the long time behaviour of this hybrid model, gives the quasi-stationary distribution. However, their work could not be generalised to a multiple type population. Campillo, Joannides and Valverde [8][9] add stochastic drift into both the nutrient and population dynamics, and provide an algorithmic method to show that the numerical results approximate the Fokker-Planck equation in a two-dimensional chemostat environment, without offering an explicit diffusion approximation.

In all the above work, no explicit derivations of the diffusion approximations for multiple species over a long time interval are given to explain the interaction between species in the same niche, and to predict the advantages or disadvantages for competition. In this thesis, by changing the time scale, we will reduce our three-dimensional hybrid system into a one dimensional diffusion approximation in order to understand the evolution of this model. Following that, dispersal (immigration) will be incorporated into the local community to achieve a explicit stationary distribution which will be compared with Hubbell's neutral model to determine the role of selection, drift, dispersal and life history trade-offs play in our local chemostat community.

Although there exist a large number of technical methods developed by mathematical microbial ecologists to calculate the diffusion approximation, to best of our knowledge, the demographic model in this thesis is the first one with an explicitly derived diffusion approximation and matched simulation results which thoroughly explains long time competition among multiple species in a resource limited local community by considering all key processes (natural selection, stochastic drift , and dispersal).

1.5 Outline of the Thesis

Chapter 1

In Chapter 1, the relevant history of mathematical population ecology and mathematical population genetics is given. We explain the differences between niche assembly theory and dispersal assembly views, present Hubbell's neutral theories and related studies that have been done in macroscopic and microscopic communities, and explain the motivations and objectives of this thesis.

Chapter 2

An introduction of the chemostat environment is first given. Besides some fundamental definitions, this chapter mainly focuses on deterministic analysis and explains the rule of niche assembly in our chemostat model. Starting from a single cell, the deterministic dynamics of the population would experience exponential and logistic growth phases before the equilibrium state is reached. Then strong selection is shown among the species, resulting in the survival of species with the largest fitness. Species sharing this largest fitness through life history trade-offs are defined as quasi-neutral. These quasi-neutral species will coexist forever in the deterministic model with its equilibrium state on a centre manifold expanded by the zero eigenvalue. With a stable assemblage of species, but no fixed equilibrium state, the equilibrium in the deterministic model is in the narrow sense, and the interaction between the quasi-neutral species is left unknown. In addition, a trade-off parameter γ_k will be defined in this chapter, whose effects on the

dynamics will be discussed later for all the models developed in the thesis.

Chapter 3

As the niche assembly theory is only able to explain the coexistence, a stochastic model is needed to predict the species diversity. The stochastic analysis is first attempted in Chapter 3, which is considered to be the most creative part of this thesis. The population density is presented as a classical birth and death process. Combined with deterministic dynamics of resource concentration, our demographic chemostat model is hybrid. The Ornstein-Uhlenbeck process then may be approximated using the law of large numbers and central limit theorem, under the condition that this is over a compact time interval. Over an infinite time interval, the law of large numbers and central limit theorem are inapplicable. The quasi-neutral species will be trapped into an absorbing state eventually, thus weak selection exists in the process. By changing the time scale, we derive a new diffusion approximation for the process over an infinite time interval. The calculation of fixation probabilities and mean of the first absorption time is a consequence of the explicit diffusion approximation. When the first absorbing state is reached, the so-called quasi-stationary distribution of the dominated species, before the whole population dies out, will be calculated.

Chapter 4

Using Monte Carlo algorithms, the stochastic model may be numerically calculated for small populations. In this chapter, all the analytical quantities derived in Chapter 2 and Chapter 3 will be compared with the results

of corresponding numerical calculations. The effects of the parameters y_k (yields), and trade-off parameter γ_k are discussed in both the phase of strong selection and weak selection. By comparison, all the analytical results match the numerical ones under a small population very well.

Chapter 5

In this chapter, the role played by the limited dispersal in shaping the local communities is mainly investigated. By introducing a small immigration rate, the deterministic dynamics and its centre manifold will not change, but the stochastic dynamic at a long time scale are no longer trapped into an absorbing state. Extinction will be balanced and coexistence will be maintained. We discuss the sensitivities of the stationary distribution of the relative abundance to the immigration rates, yields and γ_k, and state the departure from Hubbell's neutral model when the immigration rate is small.

Chapter 6

This chapter, besides summarising what has been done in the thesis, also identifies some open tasks associated with this work and remaining to be done.

Chapter 2

The Chemostat and Its Deterministic Model

The open microbial community this thesis considers is one which excludes predator-prey, parasite-host and mutualistic relations, and includes only species that pursue similar ways of life and compete with each another for the same nutrients. A chemostat (from **chem**ical environment is **stat**ic) is a classical environment for this scenario, and an important research technique in microbiology and population biology.

This chapter will give an idea of what the chemostat environment is and how its deterministic model works. In the first section, the chemostat environment will be clearly defined with essential assumptions. After giving the expressions of stochastic and deterministic processes for both population density and nutrient concentration in Section 3.2, we will use the rest of this chapter to investigate the details of the deterministic dynamics. With the

help of Hubbell's analytical results [33] for competition between non-neutral species, the results of strong selection and hence the niche assembly rule will be derived in Section 3.3. At the same time, the ideas of fitness, quasi-neutrality and life history trade-offs will be defined. The specific case of two quasi-neutral species and its equilibrium states will be explained in Section 3.4. In the last section before the summary, we will introduce a trade-off parameter γ_k which is important throughout this thesis.

2.1 Description of the Chemostat

2.1.1 Introduction to the Chemostat

Our chemostat is a bioreactor as shown in Figure 2.1. It contains a population of n bacterial species, each with density $x_k(t) = \frac{N_k(t)}{V}$ (number of cells at time t/unit volume, $k = 1..n$), and a single limited substrate of concentration $S(t)$ to feed the bacteria in a culture vessel (it is assumed that all other substrates are supplied in excess of demand). Due to resource limitation, the population in this closed volume is kept finite.

As Figure 2.1 shows, the chemostat environment is continually supplied with stock nutrient of concentration S^{in} from a nutrient reservoir with a constant flow at rate D (in units 1/time). At the same time, the culture liquid is continuous removed with the same dilution rate D to keep the culture volume constant. By changing the rate of inflow, the growth rate of the bacteria can be easily controlled. The bioreactor is assumed to be sufficiently well stirred, so that the bacteria and nutrient are spatially uniformly distributed, and each individual bacterium has equal access to the nutrients.

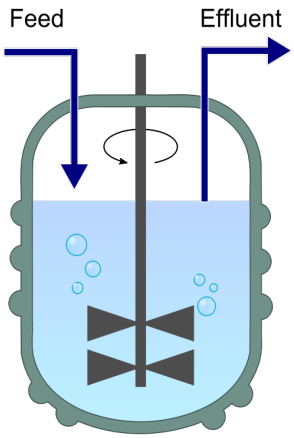

Figure 2.1: A completely stirred chemostat bioreactor with a continuous influx (the feed) and outflow (the effulent).

The chemostat environment is perhaps the best laboratory idealisation of nature for competitive population studies. All such models are all open systems for energy and materials. The input and removal of nutrients to and from the chemostat represents the continuous turnover of nutrients in nature. The outflow of organisms is formally equivalent to nonspecific death, predation, or emigration, which always occur in nature [10].

2.1.2 Monod Function: Function for the Growth Rate of Bacteria

Defined as an increase in the number of cells, growth is an essential component of microbial function. The growth rate describes the change in cell number per unit time. The mathematical expression for the growth rate has been studied over the centuries. The logistic equation Eq. (1.1) is one of the most widely used ones, which expresses the rate of reproduction as

proportional to both the existing population and the amount of available resources. However, in a chemostat environment, the growth curve of bacterial may not be logistic during some initial period.[79] We need a more accurate function to explain the mechanism for the deceleration of the bacterial growth rate, namely nutrient depletion. By performing experiments on bacteria fed a single limited nutrient, a function which fitted the data best was found by Jacques Monod to describe its growth rate $b(t)$. It is called the Monod function, with a resource $S(t)$ dependent expression:

$$\text{Monod Function:} \quad b(S) = \frac{mS(t)}{a + S(t)} \ .$$

As illustrated in Figure 2.2, the growth function monotonically increases, approaching the maximum limit value m as $S \to \infty$. When S reaches the value a (in units mass/volume), the value of growth rate equals half the value of the maximum growth rate, i.e., $m/2$ (in units 1/time). Therefore m and a are called the maximum growth rate and half-saturation constant respectively. Monod also found that the rate of nutrient consumed by bacteria was proportional to the rate of bacterial growth, i.e., a gain of one unit of bacteria requires $\frac{1}{y}$ units of nutrient. y is then defined as the growth yield, and expressed as

$$y = \frac{\text{cells of bacteria produced}}{\text{mass of nutrient used}}.$$

Given these definitions, the deterministic and stochastic dynamics of species population and nutrient in chemostat may be developed in detail.

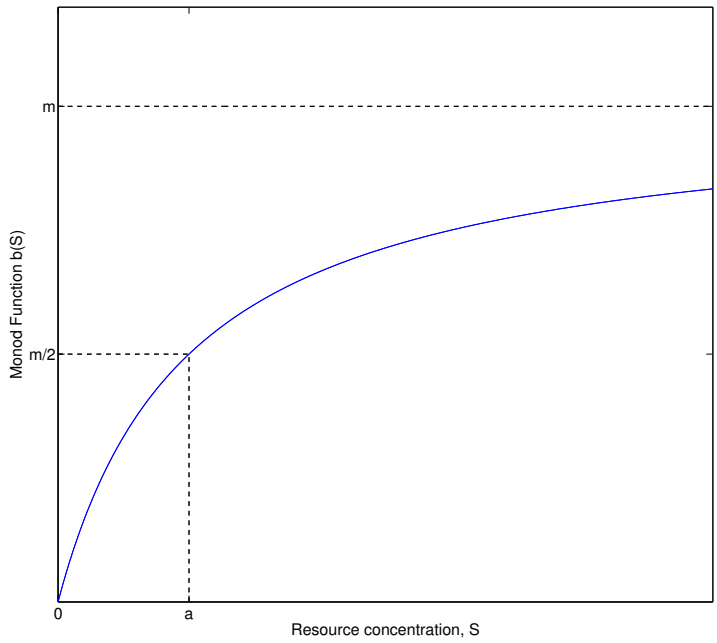

Figure 2.2: Monod Function $b(S) = \frac{mS}{a+S}$ for a species with parameter a as the half-saturation constant and m as the maximum growth rate.

2.2 Dynamics of the Chemostat Model

Before presenting our deterministic model, some essential concepts and expressions used throughout this thesis will be given.

2.2.1 Deterministic and Stochastic Dynamics

As introduced in Chapter 1, different views in community ecology arise by putting different emphases on natural selection and ecological drift. In mathematical biology, different behaviours are yielded from different approaches by treating the dynamics in discrete or continuum, deterministi-

cally or stochastically [36].

Continuous Deterministic Dynamics

Modelling begins typically in the form of ODESs in the continuous deterministic setting.

As the nature of the microbiological world is such that a ten gram sample of soil can contain as many as 10^9 individual microorganisms [63], the whole population in our chemostat model $N(t)$ is ideally close to infinity. Correspondingly, in our analytical models, by measuring in the units of volume of a typical cell, the value of volume V is also close to infinity. In this manner, a macroscopic description for the time evolutions of population density $\frac{N(t)}{V}$ and substrate concentration $S(t)$ can be given in the form of ODEs.

Stochastic Dynamics

In contrast, since noise cannot be completely ignored in the real world, the stochastic model is more suitable to describe the real dynamics. In a stochastic model, drift is present and the states of the population are variable and need to be described by probability distributions instead of unique values.

Except this chapter, all the chapters of this thesis will focus on the stochastic analysis to demonstrate how stochastic drift will shape the structures of the local community. In continuous and discrete settings for different population scales, we will develop analytical and numerical models in the later chapters, and such comparisons will be used to interpret and verify the stochastic model.

Although stochastic drift is important, the deterministic approach to ecological dynamics has often been taken, due to its straightforwardness and ability to explain natural selection and state a clear direction of evolution. Moreover, the usual approach in most studies to derive the time dependent stochastic approximation starts from expanding the deterministic ODEs. Therefore, to give some necessary insights, the deterministic model will be first discussed in the rest of this chapter .

2.2.2 Dynamics of the Population Density and Substrate Concentration

The dynamics of the chemostat model in this thesis are compromised of the population densities and substrate concentration. Thanks to Jacques Monod's work, their interaction is precisely defined, as stated in Subsection 2.1.2. In this subsection, the description of the dynamics for the substrate concentration and population densities in both deterministic and stochastic models will be given. In this thesis, the substrate concentration is measured in units of mass/volume and the species density in units of cells volume.

Dynamics of the Population Densities $x_k(t)$

From our previous assumptions, the birth rate for each species is expressed as a substrate-dependent Monod function, $b_k(t) = \frac{m_k S(t)}{a_k + S(t)}$. The death rate for each species is defined as $d_k = D + \mu_k$, with dilution rate of the liquid D and the species intrinsic death rate μ_k. We assume the value of dilution rate D in this thesis to be much larger than the intrinsic death rate μ_k. Therefore in the later analytical analysis, μ_k is insignificant to impact

the results.

Since in the deterministic dynamics, all the birth and death events occur deterministically, the density of species type k can be expressed in ODE form:

$$\frac{\mathrm{d}x_k}{\mathrm{d}t} = \frac{m_k S(t)}{a_k + S(t)} x_k(t) - (D + \mu_k) x_k(t). \tag{2.1}$$

In the stochastic dynamics, the population is assumed to behave as a counting processes with Markovian rates. We will then explain in Chapter 3 how the stochastic process can be developed by adding stochastic drift into the deterministic dynamical equation Eq. (2.1).

Dynamics of the Substrate Concentration $S(t)$

Assume the substrate concentration in the input flow S^{in} is fixed. With dilution rate of the culture liquid D (which describes the fraction of volume being replaced in a unit of time), the input rate of nutrient in the local environment is $S^{\text{in}}D$. The removal of nutrient is composed of two parts: one is removed by dilution at rate $S(t) \cdot D$, the other part is used by the growing bacteria,

$$\frac{b_k N_k}{y_k} = \frac{\text{rate of production of bacteria}}{\text{number of bacteria produced per unit substrate}},$$

where b_k is the growth rate of the k^{th} bacterial species.

In both deterministic and stochastic formulations of the model, the rates of change of substrate concentration are defined to be deterministic and can

be defined by the following ODE:

$$\begin{aligned}\frac{\mathrm{d}S}{\mathrm{d}t} &= S^{\mathrm{in}}D - \Big(S(t)D + \sum_{k=1}^{n}\frac{1}{V}\frac{b_k N_k}{y_k}\Big)\\ &= \Big(S^{\mathrm{in}} - S(t)\Big)D - \sum_{k=1}^{n}\frac{x_k(t)}{y_k}\frac{m_k S(t)}{a_k + S(t)}.\end{aligned}$$

Using the above descriptions, in the rest of this chapter we will derive the deterministic model, and we will study the stochastic model in the following chapters. To distinguish them, we will use $(\bar{S}(t), \bar{x}(t))$ to express the deterministic process, and $(\hat{S}(t), \hat{x}(t))$ for the stochastic process.

2.3 Deterministic Model

With the descriptions and assumptions in Section 3.2, the deterministic process

$$\Big(\bar{S}(t), \bar{x}(t)\Big) = \Big(\bar{S}(t), \bar{x}_1(t), \ldots, \bar{x}_n(t)\Big),$$

may be expressed as the solution of the ODE system:

$$\frac{\mathrm{d}\bar{S}}{\mathrm{d}t} = \Big(S^{\mathrm{in}} - \bar{S}(t)\Big)D - \sum_{k=1}^{n}\frac{\bar{x}_k(t)}{y_k}\frac{m_k\bar{S}(t)}{a_k + \bar{S}(t)}, \tag{2.2}$$

$$\frac{\mathrm{d}\bar{x}_k}{\mathrm{d}t} = \frac{m_k\bar{S}(t)}{a_k + \bar{S}(t)}\bar{x}_k(t) - (D + \mu_k)\bar{x}_k(t), \tag{2.3}$$

$$\bar{S}(0) = S_0 > 0,$$

$$\bar{x}_k(0) = x_{k,0} > 0. \qquad k = 1, 2, \ldots, n$$

The deterministic dynamics with three species are plotted in Figure 2.3 with their birth rate (Monod) functions.

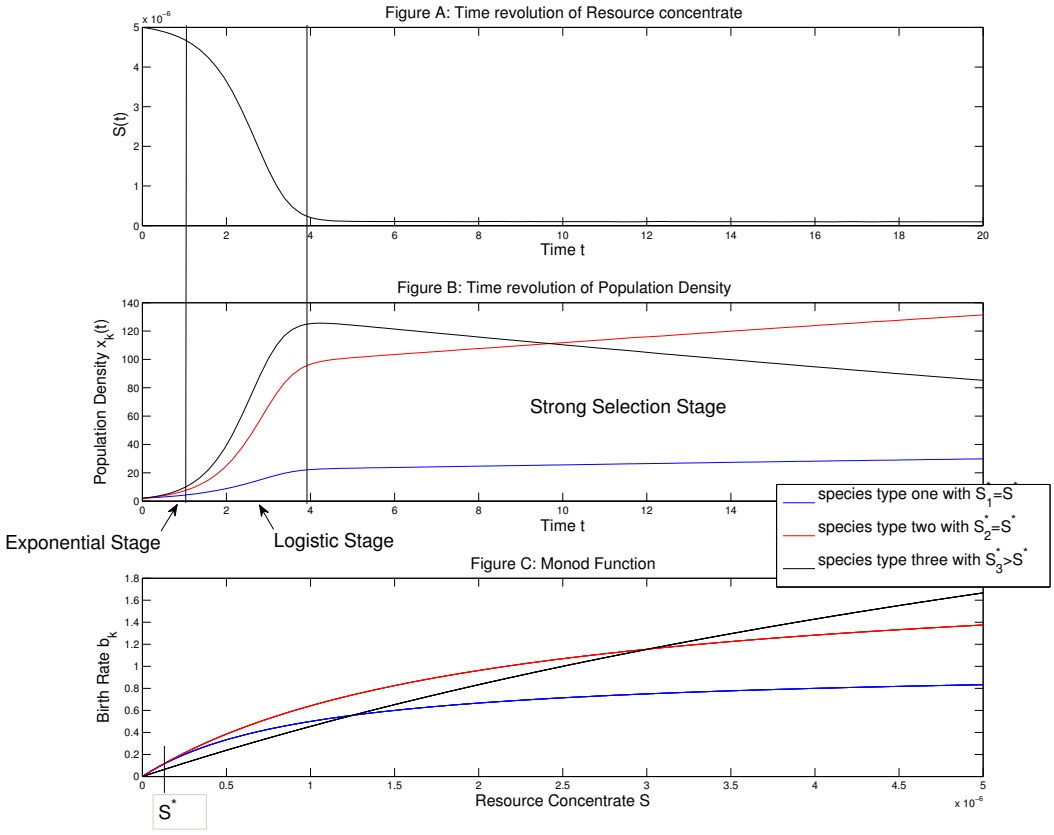

Figure 2.3: This figure illustrates the deterministic dynamics of the chemostat model with three species. Figure A and B are the time evolutions of the substrate concentration and species densities. Figure C plots the growth rate functions (Monod functions) at different resource concentrate value for each species. The blue lines in Figure B and C represent the first species with parameters: $a_1 = 1 \cdot 10^{-6}$, $m_1 = 1$, $y_1 = 5 \cdot 10^7$. The red lines are the second species with: $a_1 = 2 \cdot 10^{-6}$, $m_1 = 1.925$, $y_1 = 5 \cdot 10^7$. The black lines give the third species with: $a_1 = 1 \cdot 10^{-5}$, $m_1 = 5$, $y_1 = 5 \cdot 10^7$.

Let us define S_k^* as the steady substrate concentration when the kth species is grown alone in the chemostat. By setting Eq. (2.3) to be zero, $S_k^* = \frac{a_k(D+\mu_k)}{m_k-(D+\mu_k)}$ is yielded for each species.

2.3.1 Overview of the Deterministic Dynamics

As shown in Figure 2.3 A, the substrate is consumed dramatically at the beginning by the increasing population and then reaches a low equilibrium level S^* to establish the carrying capacity of the community. Correspondingly, in Figure 2.3 B, the population has enough resource to feed itself at first, but competition among species for the limited nutrient starts once the carrying capacity is reached. In Figure 2.3 B, the deterministic time evolutions of population densities can be separated into an exponential phase, a logistic phase and a strong selection phase. These phases will be explained in the following Subsection 2.3.2. Figure 2.3 C helps to explain how the species with different birth rate coefficients will behave in the deterministic dynamics. The birth rate (Monod) function is monotonic with substrate concentration S. As the substrate concentration reduces, the birth rate will decrease correspondingly to a stable level where it equals the death rate and then net growth vanishes.

By comparing Figure 2.3 C with Figure 2.3 B, it can be seen that the third species with the largest value of S_k^* (in black lines) has the smallest birth rate at S^* (in Figure C), and will die out eventually (in Figure B). The first and second species sharing the same value of $S_k = S^*$ will coexist for a long time. The second species with the larger slope of its birth rate at S^* (in red lines) easily reaches a higher proportion in the population. This selection will be discussed in Subsection 2.3.2, and the behaviour of the surviving species will be analysed in Subsections 2.3.3 and 2.3.4.

2.3.2 Phases in the Time Evolution

By assuming that each species starts from a single cell and $S_0 = S^{\text{in}}$, in this subsection, we will discuss the behaviour of the deterministic dynamics in the different time phases.

2.3.2.1 Phase 1: Exponential Growth Phase

At very beginning, the resource is sufficient, and since it has a very small initial value, the population will not reach a level to impact the substrate concentration. The substrate concentration will stay at S^{in}, and the dynamics may be approximated by,

$$\begin{aligned}
\bar{S}(t) &= S^{\text{in}}, \\
\frac{\mathrm{d}\bar{x}_k}{\mathrm{d}t} &= \frac{m_k \bar{S}(t)}{a_k + \bar{S}(t)} \bar{x}_k(t) - (D + \mu_k)\bar{x}_k(t), \\
\bar{x}_k(0) &= x_{k,0} \qquad\qquad k = 1, 2, \ldots, n
\end{aligned}$$

with solution,

$$\begin{aligned}
\bar{S}(t) &= S^{\text{in}}, \\
\bar{x}_k(t) &= x_{k,0} e^{\left(b_k S^{\text{in}} - D - \mu_k\right) t}.
\end{aligned}$$

As the initial value of the species density, $x_{k,0}$, is very small, by taking the log of $\bar{x}_k(t)$, we obtain

$$\begin{aligned} \bar{S}(t) &= S^{\text{in}} \\ \log \bar{x}_k(t) &= \Big(b_k(S^{\text{in}}) - D - \mu_k\Big)t. \end{aligned} \tag{2.4}$$

This expression shows that the species experiences exponential growth of its density, with constant rate $b_k(S^{\text{in}}) - D - \mu_k$. We define this phase as the exponential growth phase.

Since the intrinsic death rate is small enough to be negligible compared to the dilution rate, the death rate $D + \mu_k$ can be approximated as D. Then from the exponential equation Eq. (2.4), the species with the higher birth rate in the beginning, $b_k(S^{\text{in}})$, will dominate in this phase. However, due to the dramatic growth of the population, the resource concentration will rapidly deplete and the logistic growth phase is soon reached.

2.3.2.2 Phase 2: Logistic Growth Phase

After a small period of time in which the population grows exponentially, the population reaches $N_k(t) = O(N)$[1], where the parameter N determines the typical population size, and the substrate consumed by the population may no longer be neglected. There is a sharp drop in substrate concentration as shown in Figure 2.3 A. A decrease in birth rate follows this drop, but it is not affected sufficiently to stop the growth of the population which keeps accelerating until it approaches the middle point of this phase. As the

[1] If $f(N) = O(g(N))$, then $\lim_{N\to\infty} \sup |\frac{f(N)}{g(N)}| < \infty$

resource level continues to reduce, it will reach a point where the birth rate b_k decreases to a level close to that of the death rate d_k which equals $D+\mu_k$. Expressions for the deterministic process in this period are,

$$\frac{\mathrm{d}\bar{S}}{\mathrm{d}t} = \Big(S^{\mathrm{in}} - \bar{S}(t)\Big)D - \sum_{k=1}^{n} \frac{\bar{x}_k(t)}{y_k} \frac{m_k \bar{S}(t)}{a_k + \bar{S}(t)} < 0$$
$$\frac{\mathrm{d}\bar{x}_k}{\mathrm{d}t} = \Big(\frac{m_k a_k}{a_k + \bar{S}(t)} - (D+\mu_k)\Big)\bar{x}_k(t) > 0.$$

Since the solution of this system is in the form of a sigmoid curve (a logistic curve which is S shaped), we will call this time period the logistic phase.

2.3.2.3 Phase 3: Strong Selection Phase

When the growth of the population meets resistance due to substrate limitation, competition begins. The niche assembly rules explains that, competition among species is the ecologically equivalent of selection among genotypes, and is expected to have the same outcome at equilibrium: the best-adapted species will have replaced all others [6]. In this subsection, we will explain this selection process in our chemostat model analytically.

At equilibrium the time derivatives of the state variables equal zero [40]. Using Eq. (2.3), the species type k surviving in equilibrium state ($\bar{x}_k > 0$) then satisfies

$$0 = \frac{m_k S_k^*}{a_k + S_k^*} - d_k = \frac{m_k S_k^*}{a_k + S_k^*} - \mu_k - D.$$

Different types of species with different intrinsic parameters may result in different S_k^*, However, these species cannot coexist as there is only a single

final steady substrate concentration S^*.

Using the analysis in Hsu et al's 1977 paper [33], we will explain that the species with the smallest parameter $S_k^* = \frac{a_k}{\frac{m_k}{D+\mu_k}-1}$ will outcompete all other species in the chemostat environment. We name this deterministic phase the strong selection phase, because of its deterministic property when compared to the weak selection in the stochastic model.

Given the parameters of the chemostat model with one limited nutrient (growth rates, Michaelis-Menten constants, input concentration of the limiting nutrient, and dilution rates), Hsu et al [33] analytically explain the deterministic competition among non-neutral species. By focusing on the parameter $S_k^* = \frac{a_k}{\frac{m_k}{D}-1}$, Hsu et al prove mathematically that only the species with the smallest value of S_k^* survives, no matter how abundant the competitors are at the start, or how efficiently the species convert the substrate into cell growth (yields y_k) [33].

We will keep most of the definitions in Hsu et al's chemostat model except those of death rates and the number of species surviving. By introducing an intrinsic death rate μ_k into Hsu et al's model, we relabel the species to make sure their parameters $S_k^* = \frac{a_k}{\frac{mk}{D+\mu_k}-1}$ are ordered as

$$0 < \frac{a_1}{\frac{m_1}{D+\mu_1}-1} < \frac{a_2}{\frac{m_2}{D+\mu_2}-1} \leq \cdots \leq \frac{a_n}{\frac{m_n}{D+\mu_n}-1}.$$

With at least $\frac{a_1}{\frac{m_1}{D+\mu_1}-1} < S^{\text{in}}$, otherwise no species will survive.

By a straightforward generalisation of Hsu et al's results [33], first species has the lowest value of S^* so it survives and outcompetes all other species.

The limiting values are

$$\lim_{t\to\infty} \bar{x}_1 = \frac{y_1(S^{\text{in}} - S^*)}{1 + \frac{\mu_1}{D}},$$

$$\lim_{t\to\infty} \bar{x}_k = 0, \quad k = 2, 3, ...n,$$

$$\lim_{t\to\infty} \hat{S} = S^* = \frac{a_1}{\frac{m_1}{D+\mu_1} - 1}.$$

Therefore the parameter S_k^*, which represents the equilibrium substrate concentration when the k^{th} species is grown alone in the chemostat model, is the measurement that may be used to judge the fitness of each species. The winning species is the one with the highest fitness (lowest parameter S_k). This is the result of strong selection, and is not dependent on the cell growth yields y_k of each species.

Fitness

The parameter S_k^* forms a useful measurement for fitness in the model, as the strong selection regime selects the species with the smallest value of S_k^*. In biology, fitness describes the ability to reproduce. By defining a new parameter $\frac{1}{S_k^*}$ as the fitness of every species, we can conclude that the species with the largest fitness will survive the strong competition. Since the value of S_k^* can be measured in the chemostat with a single species grown alone, this selection outcome is deterministic, and may be predicted in advance.

2.3.3 Quasi-neutrality and Life History Trade-offs

Quasi-neutral Species

The other difference between our model and Hsu et al's chemostat model, is that there can be more than one species type sharing the same largest fitness (smallest S_k^*), i.e.,

$$0 < \frac{a_1}{\frac{m_1}{D+\mu_1}-1} = \frac{a_2}{\frac{m_2}{D+\mu_2}-1} = \cdots = \frac{a_d}{\frac{m_d}{D+\mu_d}-1} < \frac{a_{d+1}}{\frac{m_{d+1}}{D+\mu_{d+1}}-1} \leq \cdots \leq \frac{a_n}{\frac{m_n}{D+\mu_n}-1}.$$

Given the results in the strong selection phase, we know that all the species j ($j > d$) go extinct when $t \to \infty$, and species k ($k \leq d$) enter the equilibrium state,

$$\Omega = \{S^*, x^*\}$$

with elements satisfying:

$$\begin{aligned} S^* &= \frac{a_k(D+\mu_k)}{m_k-(D+\mu_k)} \\ \sum_{k=1}^{d} \frac{D+\mu_k}{y_k} x_k^* &= (S^{\text{in}} - S^*)D, \quad k = 1, \ldots, d \\ x_k^* &= 0, \quad k > d. \end{aligned} \tag{2.5}$$

These species $k = 1, \ldots, d$ which share the same largest deterministic fitness are defined as mutually quasi-neutral.

d is the number of species which survive the strong selection. In this thesis, we will assume two quasi-neutral species ($d = 2$). The situation for $d > 2$ may be similarly derived in future work based on the analysis in this

thesis.

Life History Trade-offs

In contrast to Hubbell's neutral theory which assumes that all species are competitively equivalent, we allow the quasi-neutral species to differ, through life history trade-offs:

$$\frac{a_1(D+\mu_1)}{m_1-(D+\mu_1)} = \frac{a_2(D+\mu_2)}{m_2-(D+\mu_2)} = S^*. \tag{2.6}$$

a_k, m_k and μ_k can be defined as trade-off parameters. In the later analysis, we will find that the quasi-neutral species with different trade-off parameters exhibit substantial difference in the diversity of their communities.

After the deterministic process reaches the multiple species coexistence phase, a stability problem in deterministic analysis arises: Is the equilibrium state stable? Is the coexistence state fixed? These questions will be discussed in the following subsections.

2.3.4 Equilibrium States and Their Stability Analysis

We denote the equilibrium states in our two quasi-neutral species model as $\Omega = \{S^*, x_1^*, x_2^*\}$. There are four possible stationary states in the deterministic system:

Coexistence state $\{S^*, x_1^* \neq 0, x_2^* \neq 0\}$,

Two fixation states $\{S^*, \frac{(S^{\text{in}}-S^*)Dy_1}{d_1}, 0\}$, $\{S^*, 0, \frac{(S^{\text{in}}-S^*)Dy_2}{d_2}\}$,

Extinction state $\{S^{\text{in}}, 0, 0\}$.

In the absence of immigration, when the population goes extinct it will remain extinct and, hence, the extinction state is stable.

Let us assume the strong selection phase ends in a state where both species exist. This subsection discusses the stability of the coexistence equilibrium state, and the stability of the two fixation states follows next. The first Lyapunov method [40] will be used to evaluate this stability based on an eigenvalue and eigenvector analysis.

2.3.4.1 Stability of the Coexistence Equilibrium State

In the first step, by linearising the system at the coexistence stationary point $\{S^*,\ x_k^* \neq 0\}$, we derive the Jacobian matrix $J^* = J(S^*, x_k^*)$ of the first order partial derivatives:

$$\mathbf{J}^* = \begin{pmatrix} -D - \sum_{k=1}^{d} \frac{x_k^* m_k a_k}{y_k (a_k + S^*)^2} & -\frac{d_1}{y_1} & -\frac{d_2}{y_2} \\ \frac{x_1^* m_1 a_1}{(a_1 + S^*)^2} & 0 & 0 \\ \frac{x_2^* m_2 a_2}{(a_2 + S^*)^2} & 0 & 0 \end{pmatrix} = \begin{pmatrix} -D - \sum_{k=1}^{d} \frac{(m_k - d_k)^2}{m_k a_k} \frac{x_k^*}{y_k} & -\frac{d_1}{y_1} & -\frac{d_2}{y_2} \\ \frac{(m_1 - d_1)^2}{m_1 a_1} x_1^* & 0 & 0 \\ \frac{(m_2 - d_2)^2}{m_2 a_2} x_2^* & 0 & 0 \end{pmatrix}. \tag{2.7}$$

The next step is to find the eigenvalues and eigenvectors. Denote u_1, u_2, u_3 as the eigenvectors of the Jacobian matrix J^*, and $\lambda_1, \lambda_2, \lambda_3$ as the corresponding eigenvalues. Then the deterministic dynamic system evolves as follows:

$$\bar{x}(t) = c_0 + c_1 e^{\lambda_1 t} u_1 + c_2 e^{\lambda_2 t} u_2 + c_3 e^{\lambda_3 t} u_3, \tag{2.8}$$

where $\lambda_1 = 0$ and $\lambda_{2,3}$ are the roots of

$$\lambda^2 + (D + \sum_{k=1}^{d} \frac{(m_k - d_k)^2}{m_k a_k} \frac{x_k^*}{y_k})\lambda + \sum_{k=1}^{d} x_k^* \frac{(m_k - d_k)^2}{m_k a_k} \frac{d_k}{y_k} = 0,$$

which are both negative.

The zero eigenvalue of the equilibrium state proves the existence of a centre manifold which is spanned by eigenvector u_1. The other two negative eigenvalues delineate the deterministic process tending to this centre manifold.

In the following subsection, we will expand the analysis of the centre manifold and the flows of the population dynamics toward it.

2.3.4.2 Centre Manifold and Dynamical Flows Toward It

The eigenvector of zero eigenvalue, u_1, spans a centre subspace. The centre manifold W^c is the tangent plane to this centre subspace,

$$\{(S^*, x_1^*, x_2^*) \mid S^* = \frac{a_k d_k}{m_k - d_k}\,, \frac{d_1}{y_1}x_1^* + \frac{d_2}{y_2}x_2^* = (S_0 - S^*)D,\ x_1^*, x_2^* \neq 0\}. \quad (2.9)$$

To find the flows of the dynamics towards the centre manifold, we return to the solution, which is

$$\bar{x}(t) = c_0 + c_1 e^{\lambda_1 t} u_1 + c_2 e^{\lambda_2 t} u_2 + c_3 e^{\lambda_3 t} u_3. \quad (2.10)$$

As t increases, the two negative eigenvalues $\lambda_2 = -D, \lambda_3 = -\sum \frac{(m_k - d_k)^2}{m_k a_k} \frac{x_k^*}{y_k}$ cause the solution to move towards the centre manifold in a direction that is parallel to their eigenvectors $u_2 = \begin{pmatrix} -D \\ \frac{(m_1-d_1)^2}{m_1 a_1} x_1^* \\ \frac{(m_2-d_2)^2}{m_2 a_2} x_2^* \end{pmatrix}, u_3 = \begin{pmatrix} -\sum \frac{\gamma_k}{y_k} \\ \frac{(m_1-d_1)^2}{m_1 a_1} x_1^* \\ \frac{(m_2-d_2)^2}{m_2 a_2} x_2^* \end{pmatrix}$.

For the dynamics of the substrate concentration, its equilibrium value is fixed at S^*. The flows of the dynamics of the population densities and the centre manifold are plotted in Figure 2.4, which illustrates that no matter

where the dynamics starts from, it will move back to the centre manifold.

There are no fixed equilibrium states on the centre manifold. Population processes with different initial states will intersect the centre manifold at different equilibrium points. Therefore we cannot conclude a steady state for our deterministic dynamics, but a steady number of species types (namely the number of quasi-neutral species) may be assumed.

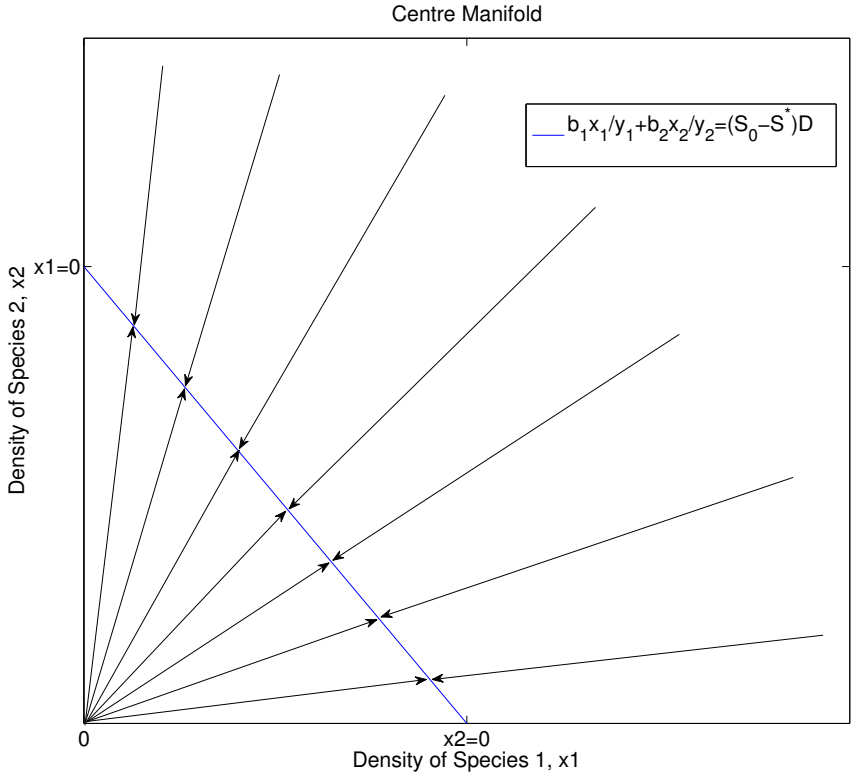

Figure 2.4: With different initial values, we numerically plotted the flow of the dynamics of the species density for parameter values: $m_1 = 1$, $m_2 = 0.9$, $a_1 = 1 \cdot 10^{-6}$, $1.925 \cdot 10^{-6}$, $y_1 = y_2 = 5 \cdot 10^8$, $D = 0.075$. The blue line gives the centre manifold, $b_1 \frac{x_1}{y_1} + b_2 \frac{x_2}{y_2} = (S^{\text{in}} - S^*)D$, where $b_k = \frac{m_k S^*}{a_k + S^*}$.

2.3.4.3 Stability of the Single-Species Equilibrium State

Referring to the equilibrium states with only single species left, namely, the two boundary states on the centre manifold, the Jacobian matrix is

$$\mathbf{J}^* = \begin{pmatrix} -D - \frac{(m_k - d_k)^2}{m_k a_k} \frac{x_k^*}{y_k} & -\frac{d_k}{y_k} \\ \frac{(m_k - d_k)^2}{m_k a_k} x_k^* & 0 \end{pmatrix}.$$

It has two negative eigenvalues

$$\lambda_1 = -D \ \ , \ \lambda_2 = -\frac{(m_k - d_k)^2}{m_k a_k} \frac{x_k^*}{y_k}.$$

Therefore, if there is only one species fixed in the population, there exists a fixed stable equilibrium state, which is $\left(S^* = \frac{a_k d_k}{m_k - d_k}, x_k^* = y_k(S^{\text{in}} - S^*)\right)$.

2.4 Parameters and Their Effects on the Deterministic Dynamics

In our chemostat model, there are four intrinsic parameters for each species, the yields (y_k), the two birth rate parameters (a_k, m_k), and the intrinsic death rate (μ_k). The later three parameters are are also called as trade-off parameters. Let us now assume the intrinsic death rate to be small enough ($\mu_k \approx 0$), and that all other species with lower fitness hence died out already, so that only the 2 quasi-neutral species remain.

As discussed in Subsection 2.3.2.3, we know that the results for the strong selection are not dependent on the yields y_k, but they do depend on the trade-off parameters (m_k and a_k). In this section, we will discuss how

these trade-off parameters and yields y_k of the quasi-neutral species will affect the long time equilibrium state.

2.4.1 Trade-off Parameters

Since the trade-off parameters m_k and a_k are also birth rate parameters, firstly we will discuss how the trade-off parameters will affect the dynamics through the birth rates.

We plot the birth rates (Monod function: $b_k = \frac{m_k S}{a_k+S}$) of the two quasi-neutral species in Figure 2.5. The two birth rates are both monotonic with substrate concentration S, and meet once at the equilibrium, where $b_1(S^*) = b_2(S^*) = D$. Therefore, in both the exponential growth phase and the logistic growth phase, where the substrate concentrations are far beyond the equilibrium level, i.e., $S(t) > S^*$, the species with the advantage of larger birth rate grows faster till the centre manifold is reached, where the net growth rates equal zero.

This advantage may be predicted by the first derivative of the birth rate with respect to the resource at the equilibrium state,

$$\gamma_k = (\frac{\mathrm{d}b_k(S)}{\mathrm{d}S})|_{S^*} = \frac{a_k}{m_k}\frac{D^2}{S^{*2}} = \frac{(m_k - D)^2}{m_k a_k} = \frac{D}{S^*}(1 - \frac{D}{m_k}). \tag{2.11}$$

This dimensionless parameter γ_k measures the responsiveness of the microbial growth rate to changes in the substrate concentration at the equilibrium value, and is plotted in Figure 2.5 as the tangent lines of the Monod functions at the equilibrium state. It can been seen that the species with large γ_k has larger birth rate where $S(t) > S^*$.

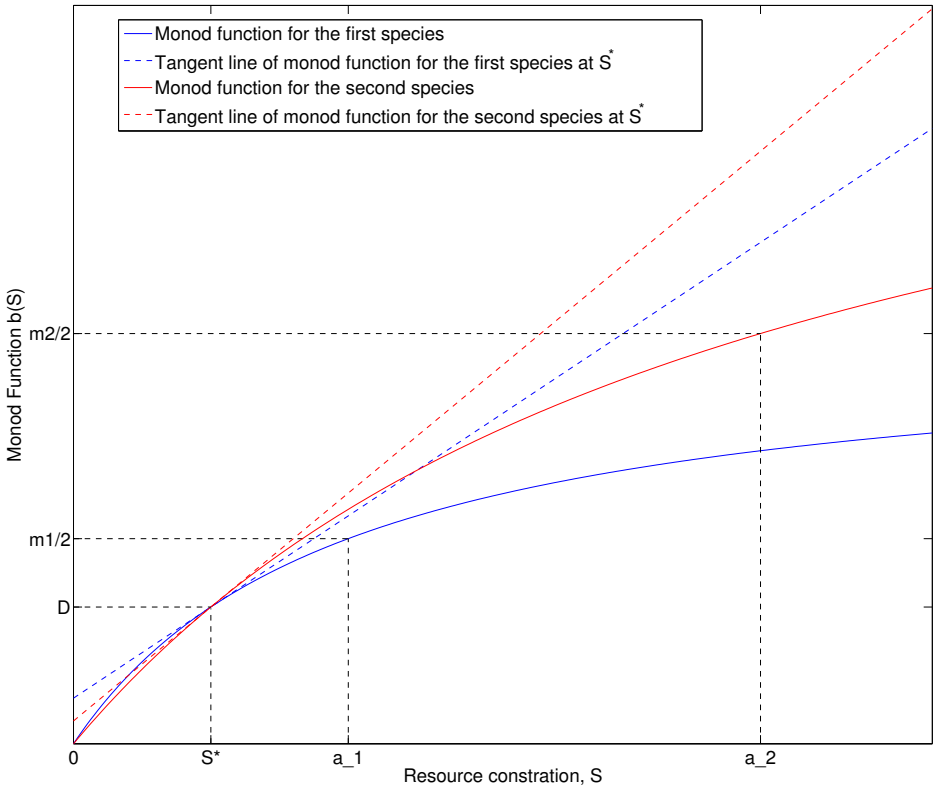

Figure 2.5: Monod functions and their derivatives for two quasi-neutral species. The blue line represents the Monod function of the species with $m = 1$, $a = 80$, blue dots are its tangent line at equilibrium state where $S^* = 40$. The red line and dots correspond to the Monod function and its tangent line for species with parameters: $m = 2$, $a = 200$ with same equilibrium value $S^* = 40$.

Since γ_k is dependent on m_k and a_k only, we denote a trade-off parameter together with m_k and a_k. If $\gamma_1 > \gamma_2$, we yield $m_1 > m_2$, $a_1 > a_2$ from Eq. (2.11). We then define this situation as that the first species has larger trade-offs parameters. Conversely, $\gamma_1 < \gamma_2$ implies that $m_1 < m_2$, $a_1 < a_2$, and it

will be explained as that the first species has smaller trade-offs parameters. Otherwise, if $\gamma_1 = \gamma_2$, two quasi-neutral species have identical trade-offs parameters.

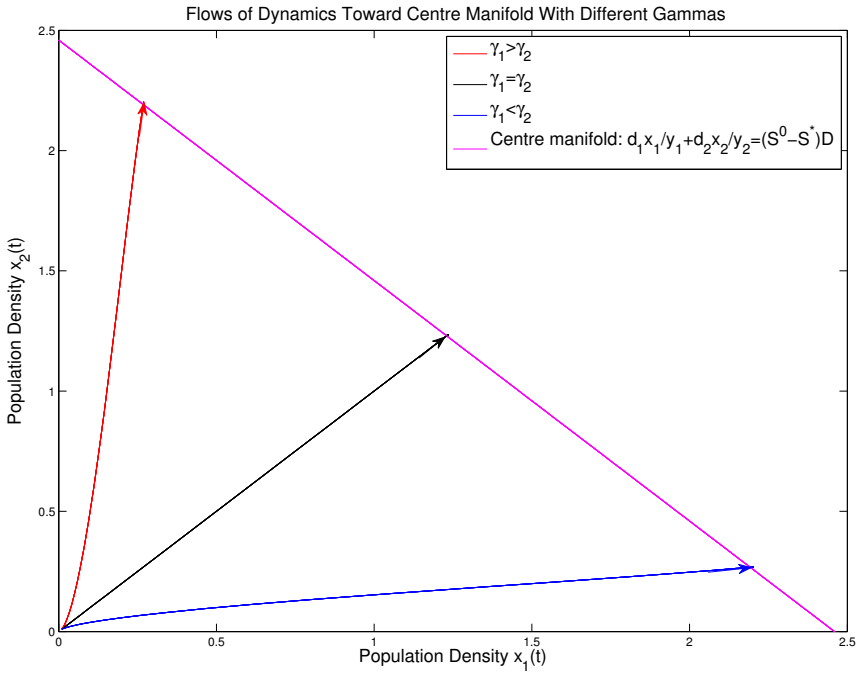

Figure 2.6: Deterministic Flows of the population dynamics with different trade-off parameter sets. The line in purple gives the centre manifold for all the models with two species sharing the same yields $y_1 = y_2 = 5 \cdot 10^5$. The lines in red, black and blue represent the flows of population with $\gamma_1 > \gamma_2$, $\gamma_1 = \gamma_2$, $\gamma_1 < \gamma_2$ respectively.

Therefore, in the quasi-neutral model, if both species start from a single cell, the species with larger trade-off parameters has an advantage in growing faster in both exponential and logistic growth phases and reaching a higher relative abundance at the equilibrium state, as illustrated in Figure 2.6.

2.4.2 Yields y_k

The growth yields y_k, are interpreted as the number of bacteria produced per unit substrate. As discussed in Section 2.3, what the yields affect is not the strong selection result and the stability of the equilibrium state, but the centre manifold with expression $\frac{d_1}{y_1}N_1^* + \frac{d_2}{y_2}N_2^* = V(S_0 - S^*)D$ and the population size $N_1 + N_2$.

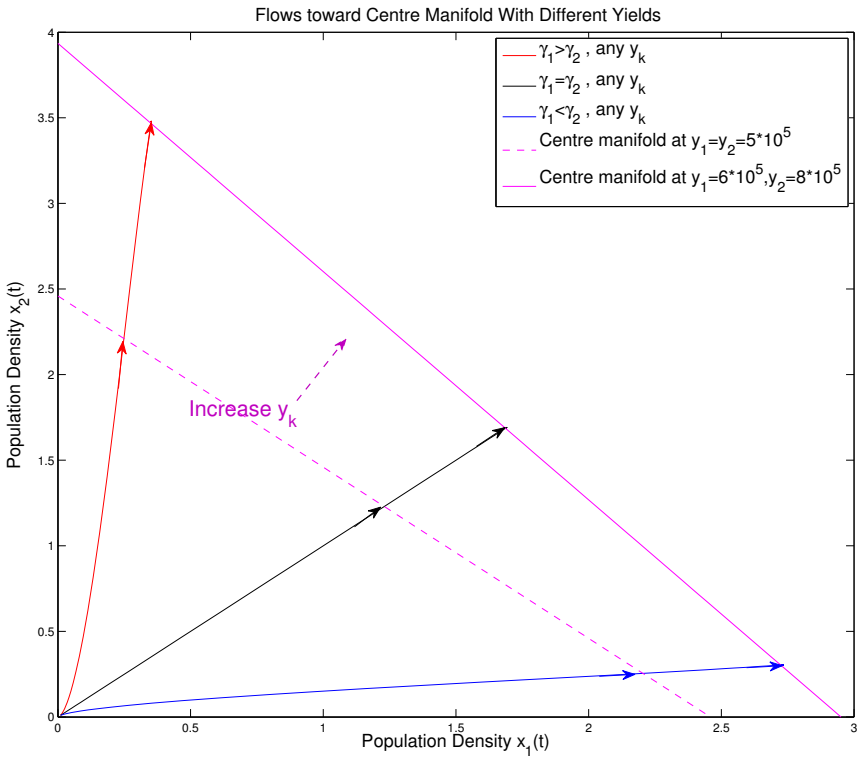

Figure 2.7: Flows of deterministic dynamics towards centre manifold with different yields. The line in purple gives the centre manifold for all the models with two species sharing the same yields $y_1 = y_2 = 5 \cdot 10^5$, and the dots in purple is the centre manifold for the models with parameters: $y_1 = 5 \cdot 10^5$, $y_2 = 8 \cdot 10^5$.

In the exponential growth phase, the species population is too small to affect the substrate concentration, therefore the effect of the yields on substrate removal may be neglected. In the following logistic growth phase, larger yields would contribute to a slower removal of the resource and consequently faster growth of the population. By plotting the flows of the population dynamics in Figure 2.7 with the same trade-offs parameters (γ_k) values in Figure 2.6, we find that upon increasing the yields, the centre manifold moves outward, and *vice versa*. However, the variations of the yields do not change the trajectories of the dynamics in the growth phases before the centre manifold is reached.

2.5 Summary and Discussion

2.5.1 Summary

In this chapter, a deterministic chemostat model is fully developed and we identify four development phases. Starting from a single cell, the population experiences the exponential growth phases and the logistic growth phases, until nutrient limitation is reached. Hsu et al analyse the strong selection regime both analytically and experimentally [33]. By applying Hsu et al's result to our model, we determine that the quasi-neutral species which share the largest fitness $\frac{1}{S^*}$ through life history trade-offs will survive under strong selection and enter the long time coexistence phase.

The zero eigenvalue in the coexistence equilibrium states spans a centre manifold. All the other eigenvalues are negative, thus the solutions of the deterministic process tend to this centre manifold. With different initial val-

ues, the dynamics will approach the centre manifold at different equilibrium states. Once it is reached, the equilibrium state will be remained. Therefore no fixed coexistence state can be predicted, but a steady-state number of species, i.e., the number of quasi-neutral species.

Lastly, we discuss the role each parameter plays in the deterministic dynamics. The fitness of a species is dependent on the trade-off parameters m_k, a_k and μ_k. If we assume a small initial condition, the species with larger trade-off parameter γ_k has the advantage of growing faster and so reaching a higher relative abundance in the coexistence population. The growth yields y_k do not affect the strong selection results and the species' growth path to the centre manifold, but will influence the time to the centre manifold and the population size.

2.5.2 Further Discussion

If we give the definition that the species in a niche share the same fitness. In this chapter, we demonstrate that niche differentiation (fitness differentiation) does allow for the coexistence of different species in the community. Niche assembly explains species survival and extinction and also promotes coexistence. However, coexistence only persists in the deterministic system if immigration and mutation are excluded. We call the species 'quasi-neutral', since the types are only neutral in terms of the deterministic limit. In the following chapters, a birth-death Markov process will be built to develop a hybrid model that incorporates both stochastic and deterministic events which may relate more closely to the natural conditions. By adding stochastic drift to the deterministic limit, the coexistence states of the quasi-neutral

species will be destroyed in the absence of immigration and mutation. The flow on the centre manifold will eventually be trapped in one of the absorbing states, where the trade-off parameters of the quasi-neutral species will show their advantage/ disadvantage. Therefore, in the long time scale, niche assembly can no longer explain the diversity of the community, but stochastic drift plays a dominant role. To balance the eventual extinction of species, dispersal-limited immigration will be incorporated in Chapter 5. Although the final coexistence steady state is still one of competitive exclusion, the time to reach a monomorphic population is then essentially indefinitely delayed [32], and the quasi-stationary distribution can be analytically and numerically measured.

In the view of Hubbell [31], niche differentiation along life-history trade-offs is the very mechanism by which per capita relative fitness is equalised among the coexisting species in a community. In this way, Hubbell accounts for both niche and dispersal limited assembly in his unified neutral theory. In the light of this, we will question if differences among the quasi- neutral species are actually important to the final diversity; and if Hubbell's unified neutral concept may be applied to our quasi-neutral chemostat model (where coexistence is maintained by the introduction of an immigration) to predict a stationary diversity. The answer to these questions will be detailed in Chapter 5.

Chapter 3

Stochastic Chemostat Model

The deterministic dynamics discussed in the previous chapter leads to the simple result that the fittest species outcompetes all others in the absence of immigration, but does not provide a stationary diversity and leaves the interaction between the quasi-neutral species poorly understood. Motivated by this fact, we shift our attention to the stochastic dynamics from this chapter onwards. Our aim is to investigate the competition by studying the impact of stochastic drift that arises due to intrinsic noise from randomness in the death/birth of individual microbes.

In real world microbial populations, a large spatial region will of course contain a relatively huge number of individuals which is close to infinity. With such a huge population, the tiny individual drift driven by demographic noises has a small effect on the deterministic limit, especially over compact time intervals. However, environmental resources are always not sufficient and – together with the saturation, competition, and prey-predator interactions – they restrict the population to be finite. Consequently the oc-

currence of demographic noise cannot be excluded, and so the deterministic model is not sufficient to describe the changes in the species frequencies and to predict the long time result. Stochastic drift then will play a significant role in the shaping of the structure in finite population.

For a small-sized population, the amplitude of the stochastic fluctuations is relatively large compared to the value of the equilibrium states, and may easily drive the whole population to the extinction state. If no immigration or mutation exists, this extinction state is absorbing. Furthermore, this occurs in not only a small population, since the limiting theory of Markov processes states that the finite state Markov chains with absorbing states will eventually be trapped into one of the absorbing states as time passes. Therefore, during this process, the deterministic stable assemblage of coexisting species (quasi-neutral species) is broken and a new competitive dynamics occurs amongst the quasi-neutral species after the process approaches the deterministic equilibrium state.

In the resource-limited, immigration-absent chemostat model there exists one final absorption state: the extinction of all species. Nevertheless, this absorption time will be far beyond the human time scale. The first absorption state where the population is monotypic will be reached long before the extinction of all species. Without the presence of mutation and immigration, once this state is achieved, it will remain until the final absorption. We call the selection for quasi-neutral species to survive the first absorption 'weak selection'. A number of pertinent questions on quantity arise at this phase. With what probability will certain species fade away? How long it will take certain species to dominate the whole population?

How sensitive are these quantities of probability and time to the absorption to the intrinsic parameters of the species and their initial densities? What is the trajectory of the stochastic process along the centre manifold before it reaches the first absorption state? To answer these questions, it is necessary to build a proper stochastic model and to analyse its dynamics.

3.1 Introduction and Overview

3.1.1 Introduction

In the first Section, a proper stochastic model will be built and defined. By applying the law of large numbers in Section 3.3, this stochastic process will be shown to converge to its corresponding deterministic limit in the limit of infinite V.

After these two fundamental sections, due to the weakness of the law of large numbers in the finite V, the analysis on stochastic dynamics is divided into two parts: dynamics over a compact time interval (Section 3.4) and dynamics over an infinite time scale (Sections 3.5, 3.6 and 3.7). Over a compact time interval, with the help of the central limit theorem, the diffusion approximation will be derived as an Ornstein-Uhlenbeck process fluctuating around the equilibrium state where the quasi-neutral species coexist. Over an infinite time interval, this coexistence will however not be maintained, instead, the population will be trapped into one of the absorbing states. This clearly highlights the significance of stochastic drift as a mechanism affecting the dynamics over long time scales. Hence, by changing the time scale, it is necessary to derive a new diffusion process in order to explain

the stochastic dynamics over an infinite time interval – which will be the main content of Sections 3.5 and 3.6. In Section 3.5, a mathematical trick will be devised. By projecting the stochastic process onto its deterministic trajectory as $t \to \infty$, we find that the trajectory on the projection map can well approximate the stochastic process. The derivatives of this projection map with respect to the initial states are calculated in this section. Later, by Itô's transformation and the weak convergence, the diffusion approximation of the projection map will be analytically derived in Section 3.6. The dynamics of the quasi-neutral species on the centre manifold changes the splitting probability which is also known as the fixation probability. There are two possible ways to calculate this result: by solving the splitting probability on the centre manifold between two absorbing points; or by using the backward Kolmogorov equation. The latter one was chosen to calculate the fixation probability and the mean of the first absorption time, which is shown in Section 3.7. Once the first absorbing state is reached, the population will spend a long time in this phase, thus it is necessary to discuss the quasi-stationary distribution of this single species and that is done in Section 3.8. In the same section, the conditional probability density of the relative abundance will be proved not to be stationary distributed. Section 3.9 concludes the whole chapter.

3.1.2 Overview of the Stochastic Dynamics over Time

Before attempting to extend the development of the stochastic model, it is useful to outline the development of the stochastic dynamics over time, to see how it differs from the deterministic dynamics of the chemostat model

with two quasi-neutral species and immigration absent. The results of numerical calculations are illustrated in Figure 3.1 (the calculations of the numerical model will be explained in Section 4.1.1 in the next chapter), where the stochastic dynamics of three species with different intrinsic parameters are simulated.

Strong Selection and Coexistence Phase

The dynamics of the stochastic process in these two phases will be explained in Sections 3.2, 3.3 and 3.4.

In Figure 3.1, the dynamics in phase one and phase two are close to the deterministic dynamics as discussed in the previous deterministic chapter. The behaviour in phase one follows the course of strong selection with the result deterministically decided by the intrinsic parameters of the species. The species with smaller deterministic fitness (red line in Figure 3.1) will die out in this phase. Starting from a single cell population, with nutrient input, the population of the quasi-neutral species (black line and blue line) grows exponentially and then logistically at the beginning. When it comes close to the centre manifold, due to the limit of the nutrient, the mean of the net growth rate gradually decreases to zero, where the coexistence phase is reached.

Weak Selection Phase

The third time phase is the weak selection phase. The weak selection regime will be discussed in Sections 3.5, 3.6 and 3.7.

After the strong selection, the surviving quasi-neutral species enter phase

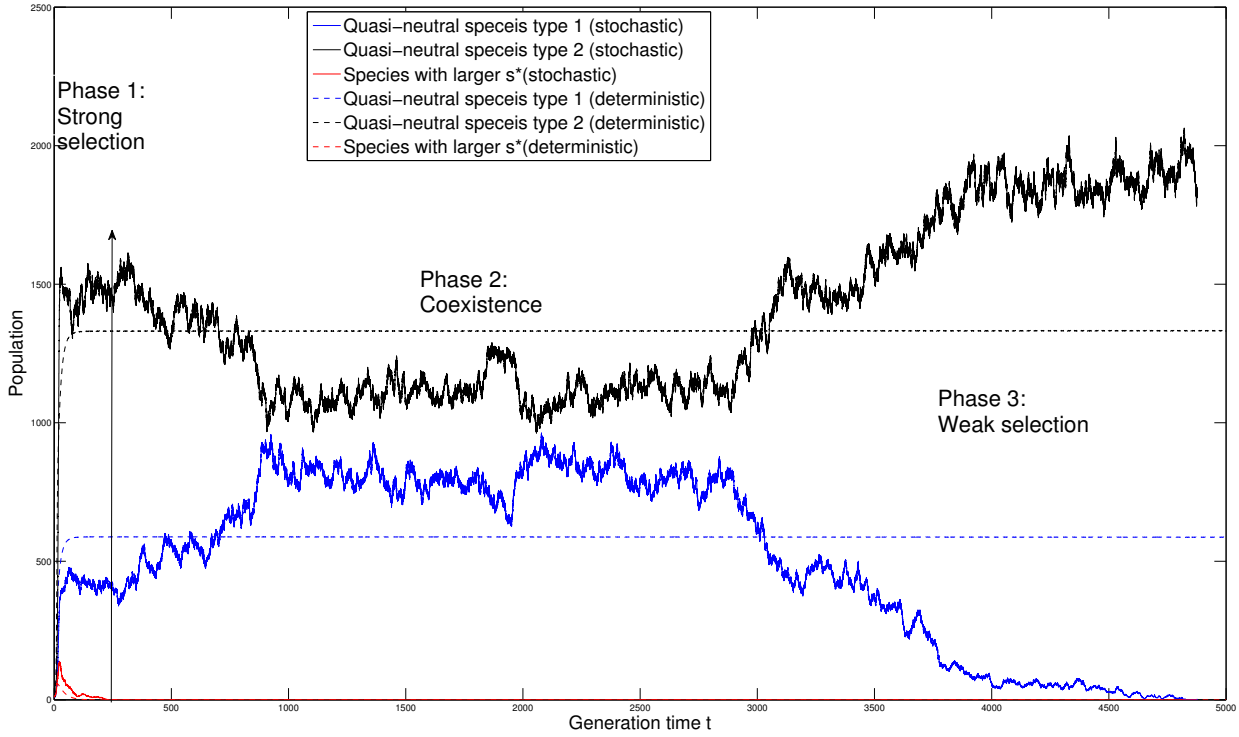

Figure 3.1: The time processes of the stochastic model with three species. All the lines are results of stochastic numerical calculations. The dots are analytical deterministic process. The red line is the stochastic process of the species with smaller fitness, where $a_{\text{red}} = 2 \cdot 10^{-6}$, $m_{\text{red}} = 1, k_{\text{red}} = 5 \cdot 10^{8}$. The red dots is the corresponding deterministic process. The stochastic process of the two quasi-neutral species (in blue and black line) in phase 1 and phase 2 fluctuate around their deterministic process (in blue and black dots), with $a_{\text{black}} = 1 \cdot 10^{-6}$, $m_{\text{black}} = 1$, $y_{\text{black}} = 5 \cdot 10^{8}$, and $a_{\text{blue}} = 2 \cdot 10^{-6}$, $m_{\text{blue}} = 1.925$, $y_{\text{blue}} = 5 \cdot 10^{8}$. In phase 3, species in black line wins the weak competition and dominate the whole population from then on.

2 where they may coexist for a long time. From a deterministic viewpoint, this coexistence will be maintained indefinitely. Nevertheless, in the actual bacterial world, the noise from the environment and demography may not be

ignored. As a result of the addition of stochastic noise, there is a small drift around the deterministic limit. Since the level of drift is tiny compared to the size of the population, its influence does not become apparent very soon. From a stochastic viewpoint, in the absence of immigration, this coexistence will not be maintained indefinitely. The accumulation of small stochastic drifts forms a mechanism to eventually cause the population dynamic towards one of the absorbing states. This is illustrated in phase 3 of Figure 3.1 where one of the quasi-neutral species dies out and the other dominates the whole population. The choice of which quasi-neutral species will dominate is not predetermined, but the result of competition between two stochastic populations. Contrasting this to the deterministic strong selection, we call this competition 'weak selection'.

Quasi-Stationary Phase

After the weak selection phase, the population is dominated by a single species. Although the population cannot avoid the final extinction destination, its endurance time will be longer than the human time scale. In the phase after the weak selection and before the final extinction, the population of the surviving species behaves as a Gaussian process. We define this phase as the quasi-stationary phase. The quasi-stationary distribution of the dominating species will be calculated in Section 3.8.

3.2 Stochastic Model Description

3.2.1 Stochastic Process of the Population Densities

As defined in Chapter 2, throughout the thesis, n is the number of species in the initial state, and there are only two quasi-neutral species left after the strong selection.

In the stochastic model, all events are assumed to occur independently according to a Poisson process, and all individuals of a given species are ecologically identical. During small time interval $(t, t+\triangle t)$, the Monod function (birth rate) could be approximated to $\frac{m_k \hat{S}(t)}{a_k+\hat{S}(t)}$. Then for species type k $(k = 1, \ldots, n)$, the transition probabilities from a population of size $\hat{N}_k$ to size $\hat{N}_k + 1$ or $\hat{N}_k - 1$ at time $t + \triangle t$ are functions of $\hat{N}_k(t)$ and of substrate concentration $\hat{S}(t)$ at time t,

$$P\{\hat{N}_k(t+\triangle t) = \hat{N}_k(t) + 1\} = \frac{m_k \hat{S}(t)}{a_k + \hat{S}(t)} \hat{N}_k(t) \triangle t + o_1(\triangle t),$$
$$P\{\hat{N}_k(t+\triangle t) = \hat{N}_k(t) - 1\} = (D + \mu_k) \hat{N}_k(t) \triangle t + o_2(\triangle t).$$

By substituting the population with its density $\hat{x}(t) = \frac{\hat{N}(t)}{V}$, the transition probabilities of process $\hat{N}(t) = \Big(\hat{N}_1(t), \ldots, \hat{N}_n(t)\Big)$ can be written as,

$$P\{\hat{N}_k(t+\triangle t) = \hat{N}(t) + 1\} = \beta_{k,1}\Big(\hat{S}(t), \hat{x}(t)\Big) V \triangle t + o_1(\triangle t), \tag{3.1}$$
$$P\{\hat{N}_k(t+\triangle t) = \hat{N}(t) - 1\} = \beta_{k,-1}\Big(\hat{S}(t), \hat{x}(t)\Big) V \triangle t + o_2(\triangle t), \tag{3.2}$$

where $l \in \{1, 1-\}$ and $\beta_{k,l}$ is the jump intensity expressed as,

$$\beta_{k,1}\Big(\hat{S}(t), \hat{x}(t)\Big) = b_k(t)\hat{x}_k(t) = \frac{m_k\hat{S}(t)}{a_k + \hat{S}(t)}\hat{x}_k(t),$$
$$\beta_{k,-1}\Big(\hat{S}(t), \hat{x}(t)\Big) = d_k(t)\hat{x}_k(t) = (D + \mu_k)\hat{x}_k(t),$$

with $b_k(t)$ and $d_k(t)$ as the birth and death rates of the species k at time t.

This is a classical birth-death process, with the expected number of events at time t is

$$\int_0^t \beta_{k,l}(\hat{S}(u), \hat{x}(u))V\mathrm{d}u.$$

By theorem 4.1 in Chapter 6 of Ref [13], the Markov process $\hat{N}_k(t)$ with jump intensities Eqs. (3.1) and (3.2) can be expressed in the form of,

$$\hat{N}_k(t) = \hat{N}_k(0) + \sum_l lY_l\{V\int_0^t \beta_{k,l}(\hat{S}(u), \hat{x}(u))\mathrm{d}u\}, \tag{3.3}$$

where $Y_l(s(t))$ is an independent standard Poisson process with $s(t)$ as its expected number of events at time t. With $l = 1$ and -1, the poisson process is a pure birth and death process respectively.

In the following step, we will expand the Eq. (3.3) in detail. Define $\tilde{Y}_l(s)$=$Y_l(s) - s$ as a Poisson process centred at its expectation. The expectation of our Markov process at time t is expressed as

$$F_k(\hat{S}, \hat{N}) = \sum_l l\beta_{k,l}\Big(S(t), \hat{N}_k\Big) = \Big(\frac{m_k\hat{S}(t)}{a_k + \hat{S}(t)} - D - \mu_k\Big)\hat{N}_k(t). \tag{3.4}$$

Reaction Function F This function $F_k(\hat{S}, \hat{N})$ is usually named as the drift function in stochastic analysis. To distinguish with the conception of

stochastic drift, we denote F as the reaction function throughout this thesis.

Then the Markov process of Eq. (3.3) can be written as an integral of the deterministic reaction term and the sum of the stochastic noise terms,

$$\hat{N}_k(t) = \hat{N}_k(0) + \underbrace{\int_0^t F_k\Big(\hat{S}(u), \hat{N}(u)\Big)\mathrm{d}u}_{\text{Deterministic reaction component}} + \underbrace{\sum_l l\tilde{Y}_l\Big\{V\int_0^t \beta_{k,l}\Big(\hat{S}(u), \hat{x}(u)\Big)\mathrm{d}u\Big\}}_{\text{Stochastic noise component}}.$$

Correspondingly, the Markov process of the species density $\hat{x}(t) = (\hat{x}_1(t), \ldots, \hat{x}_n(t))$ can be expressed as the solution of,

$$\hat{x}(t) = \hat{x}(0) + \underbrace{\int_0^t F\Big(\hat{S}(u), \hat{x}(u)\Big)\mathrm{d}u}_{\text{Deterministic reaction component}} + \underbrace{\sum_l lV^{-1}\tilde{Y}_l\Big\{V\int_0^t \beta_l\Big(\hat{S}(u), \hat{x}(u)\Big)\mathrm{d}u\Big\}}_{\text{Stochastic noise component}}.$$

3.2.2 Stochastic Process of the Substrate Concentrate

Assume $\hat{S}(t)$ is the substrate concentration at time t, and $V\hat{S}(t)$ expresses the total mass of the substrate, which is augmented by a continuous substrate supply and is diminished due to the washing out of unused substrate and conversion of substrate to microbial biomass. Mass balances on the limited substrate provide us with the deterministic dynamics equation,

$$\frac{\mathrm{d}V\hat{S}(t)}{\mathrm{d}t} = V\Big(S^{\text{in}} - \hat{S}(t)\Big)D - \sum_{k=1}^{n}\hat{N}_k(t)\frac{m_k\hat{S}(t)}{y_k\Big(a_k + \hat{S}(t)\Big)}.$$

that is

$$\frac{\mathrm{d}\hat{S}(t)}{\mathrm{d}t} = F_s(\hat{S},\hat{x}) = \Big(S^{\mathrm{in}} - \hat{S}(t)\Big)D - \sum_{k=1}^{n} \hat{x}_k(t)\frac{m_k\hat{S}(t)}{y_k\Big(a_k + \hat{S}(t)\Big)},$$

where F_s is the reaction function for the substrate concentrate.

Although the rates of change of the substrate concentration are deterministic, the process of substrate concentration still behaves stochastically due to the population density dependence.

3.2.3 Hybrid Stochastic System

Finally, the expression of the stochastic hybrid chemostat system is

$$\hat{S}(t) = S(0) + \int_0^t F_s\Big(\hat{S}(u),\hat{x}(u)\Big)\mathrm{d}u \tag{3.5}$$

$$\hat{x}_k(t) = \hat{x}_k(0) + \int_0^t F_k\Big(\hat{S}(u),\hat{x}(u)\Big)\mathrm{d}u + \sum_l lV^{-1}\tilde{Y}_l\Big\{V\int_0^t \beta_{k,l}\Big(\hat{S}(u),\hat{x}(u)\Big)\mathrm{d}u\Big\} \tag{3.6}$$

where $\tilde{Y}_l(s)$=$Y_l(s) - s$ is the Poisson process centred at its expectation and $F\Big(\hat{S}(t),\hat{x}(t)\Big)$ is the system of the reaction functions with expression:

$$F_s\Big(\hat{S}(t),\hat{x}(t)\Big) = \Big(S^{\mathrm{in}} - \hat{S}(t)\Big)D - \sum_{k=1}^{n} \hat{x}_k(t)\frac{m_k\hat{S}(t)}{y_k\Big(a_k + \hat{S}(t)\Big)} \tag{3.7}$$

$$F_k\Big(\hat{S}(t),\hat{x}(t)\Big) = \Big(\frac{m_k\hat{S}(t)}{a_k + \hat{S}(t)} - D - \mu_k\Big)\hat{x}_k(t). \tag{3.8}$$

In the following, we will analysis this stochastic dynamics starting from relating it to the deterministic dynamics.

3.3 Deterministic Limit

3.3.1 The Law of Large Numbers

It is well known that the stochastic process converges to the deterministic process in the limit of infinitly large populations. The larger the population, the more precise is the convergence. The convergence in the law of large numbers has different conditions for each model with their different assumptions. The uniform law of large numbers [13] will be applied to our chemostat model in this section.

Theorem 1 (Law of Large Numbers)**.** *Suppose that for each compact* $K \subset R^n$*,*

$$\sum_l |l| sup_{x\in K} \beta_l(x) < \infty, \tag{3.9}$$

and there exists $M_k > 0$ *such that*

$$|F(a) - F(b)| \leq M_k |a - b|, \quad a, b \in K. \tag{3.10}$$

Suppose $\hat{x}$ *satisfies* (3.9)*,* $\lim_{V\to\infty} \hat{x}_k(0) = x_{k,0}$*, and* $\bar{x}$ *satisfies*

$$\bar{x}'(t) = F(\bar{x}(s)), \quad \bar{x}_k(0) = x_{k,0}, \quad t \geq 0, \tag{3.11}$$

Then for every $t \geq 0$*,*

$$\lim_{V\to\infty} \sup_{s\leq t} |\hat{x}(s) - \bar{x}(s)| = 0 \quad a.s. \tag{3.12}$$

In our stochastic model,

$$\beta_{k,1} + \beta_{k,-1} = \Big(\frac{m_k \hat{S}(t)}{a_k + \hat{S}(t)} + D + \mu_k\Big)\hat{x}_k(t) < m_k + D + \mu_k < \infty,$$

and

$$|F_k(a) - F_k(b)| = |\frac{m_k \hat{S}(t)}{a_k + \hat{S}(t)} - D - \mu_k||(a-b)| \leq \max\Big(|m_k - D - \mu_k|, |D + \mu_k|\Big)|a-b|,$$

which means that the conditions in Eqs. (3.9) and (3.10) are satisfied.

As a result, if $V \to \infty$, the Markov process converges almost surely to the deterministic limit, $\Big(\hat{S}(t), \hat{x}(t)\Big) \Rightarrow \Big(\bar{S}(t), \bar{x}(t)\Big)$, which is the solution of

$$\begin{aligned} \mathrm{d}\bar{S}(t) &= \Big\{(S^{\mathrm{in}} - \bar{S}(t))D - \sum_{k=1}^{n} \bar{x}_k(t) \frac{m_k \bar{S}(t)}{y_k(a_k + \bar{S}(t))}\Big\}\mathrm{d}t \\ \mathrm{d}\bar{x}_k(t) &= \Big(\frac{m_k \bar{S}(t)}{a_k + \bar{S}(t)} - D - \mu_k\Big)\bar{x}(t)\mathrm{d}t. \end{aligned}$$

3.3.2 Weakness of the Law of Large Numbers

Although this law of large numbers is a strong convergence (converges almost surely), it states the conditions under which the convergence occurs uniformly over compact time internals in the case of V close to infinity. That is to say with large but still finite V, the stochastic process is likely to be close to the deterministic process, but it still leaves open the possibility that the convergence will be violated over an infinite time interval. If the time of process is elongated enough, this possibility may occur a number of times and eventually violate the deterministic coexistence status. In the later numerical models for small populations, the convergence will easily be

broken at proper time point t.

Therefore, given this uniform law of large numbers, the stochastic process converges to its limiting deterministic process over certain compact time intervals, but not an infinite time interval. Then in the following derivations of the diffusion approximations, it will be necessary to separate the analysis into compact and infinite time intervals. Over compact time intervals, the law of large numbers works well. Starting from the same initial states, the trajectory of the deterministic dynamics in Eqs. (2.2) and (2.3) coincides with the trajectory of the deterministic limits of the stochastic dynamics. At longer times, these two trajectories separate at the time point t (t close to ∞) where the convergence breaks, then a new diffusion approximation at a long time scale will be derived.

3.4 Diffusion Approximation over Compact Time Intervals

The method to derive the diffusion approximation over compact time intervals in this section is based on the central limit theorem in Ref [13]: that with suitable re-scaling, the stochastic process converges to a centred Gaussian process with independent increments.

3.4.1 Central Limit Theorem

From Eq. (3.6), the Markov process $\hat{x}(t)$ is expressed as

$$\hat{x}(t) = \hat{x}(0) + M^V(t) + \int_0^t F\Big(\hat{x}(s)\Big)ds,$$

with martingale[1] component

$$M^V = \sum_l lV^{-1}\tilde{Y}_l\{V\int_0^t \beta_l\Big(\hat{S}(s),\hat{x}(s)\Big)ds\}, \tag{3.13}$$

where $\tilde{Y}_l$ is a Poisson process centred at its expectation. F is the reaction function defined in Eqs. (3.7) and (3.8).

Set $W_l^V(u) = V^{-\frac{1}{2}}\tilde{Y}_l(Vu)$. By the central limit theorem, we have $W_l^V \Rightarrow W_l$ (converges in distribution, as $V \to \infty$), where W_l is an independent standard Brownian motion [13]. Then the martingale component M^V in Eq. (3.13) converges to $W_l\{\int_0^t \beta_l\Big(\bar{x}(s)\Big)\mathrm{d}s\}$.

Define a new central and scaled process $\chi^V(t)$ on our stochastic process $\hat{x}(t)$ as

$$\chi^V(t) = \sqrt{V}\Big(\hat{x}(t) - \bar{x}(t)\Big) \tag{3.14}$$

where $\bar{x}(t)$ is the deterministic process. The initial condition $\chi^V(0) = \sqrt{V}(\hat{x}(0) - \bar{x}(0))$ is assumed to be sufficiently close to 0. Substitute the stochastic process $\hat{x}(t)$ Eq. (3.6) into Eq. (3.14), then $\chi^V(t)$ is expanded as,

$$\chi^V(t) = \chi^V(0)+\sum_l lW_l^V\Big\{\int_0^t \beta_l\Big(\hat{x}(s)\Big)\mathrm{d}s\Big\}+\int_0^t \sqrt{V}\Big\{F\Big(\hat{x}(s)\Big)-F\Big(\bar{x}(s)\Big)\Big\}\mathrm{d}s. \tag{3.15}$$

The purpose of applying the central limit theorem is to approximate

[1] A martingale is a stochastic process for which, at time t, the expectation of the next value in the sequence is equal to the present observed value even given knowledge of all prior observed values at the current time.

$\chi^V(t)$ in Eq. (3.15) by a new diffusion of the form of

$$X(t) = X(0) + \sum_l lW_l\Big\{\int_0^t \beta_l(\bar{x}(s))\mathrm{d}s\Big\} + \int_0^t \partial F\Big(\bar{x}(s)\Big)X(s)\mathrm{d}s, \quad X(0) = 0. \tag{3.16}$$

Then our diffusion approximation of $\hat{x}(t)$ could be yielded as

$$\hat{x}(t) \Rightarrow \bar{x}(t) + \frac{1}{\sqrt{V}}X(t).$$

Theorem 2 (Central Limit Theorem)**.** *Suppose that for each compact* $K \subset R^n$*,*

$$\sum_l |l|^2 sup_{x\in K}\beta_l(x) < \infty,$$

and that the β_l *and* ∂F *are continuous. Then* $\chi^V(t) \Rightarrow X(t)$ *(converges in distribution).*

Since all the conditions in the central limit theorem can easily be satisfied as in the law of large numbers, the process $\chi^V(t)$ in Eq. (3.15) converges to the process $X(t)$ in Eq. (3.16) in our chemostat model. Rewriting $X(t)$ in Eq. (3.16) by Theorem 5.3 in Ref [13], we have

$$X(t) = X(0) + \sum_l \int_0^t l\beta_l^{\frac{1}{2}}\Big(\bar{x}(s)\Big)\mathrm{d}W_l(s) + \int_0^t \partial F\Big(\bar{x}(s)\Big)X(s)\mathrm{d}s, \quad X(0) = 0,$$

with its stochastic differential equation (SDE) form,

$$\mathrm{d}X(t) = \partial F\Big(\bar{x}(t)\Big)X(t)\mathrm{d}t + \sum_l l\beta_l^{\frac{1}{2}}\Big(\bar{x}(s)\Big)\mathrm{d}W_l.$$

Therefore, by transforming as in Eq. (3.4.1), we derive the diffusion

approximation of our stochastic process $\hat{x}(t)$ in the form of:

$$\mathrm{d}\hat{x}(t) = \underbrace{\partial F\Big(\bar{x}(t)\Big)\Big(\hat{x}(t) - \bar{x}(t)\Big)\mathrm{d}t}_{\text{Deterministic term}} + V^{-\frac{1}{2}} \sum_l l\beta_l^{\frac{1}{2}}\Big(\bar{x}(s)\Big)\mathrm{d}W_l. \tag{3.17}$$

where $\bar{x}(t)$ is the deterministic limit of stochastic process, the deterministic term is linearised, and F is the reaction function expressed as in Eq. (3.4). This approximation will be expressed in detail in the following subsection.

3.4.2 Diffusion Approximation

From the previous deterministic chapter, we know that while the centre manifold is approached, $\bar{x}(t)$ will keep on it in definitely. Replace $\bar{x}(t)$ in Eq. (3.17) by the equilibrium state $x^* = (S^*, x_1^*, x_2^*)$, then the diffusion approximation of our stochastic process $\hat{x}(t)$ is obtained of the stochastic differential equation (SDE) form,

$$\mathrm{d}\hat{x}(t) = \partial F(x^*)\Big(\hat{x}(t) - x^*\Big)\mathrm{d}t + V^{-\frac{1}{2}} \sum_l l\beta_l^{\frac{1}{2}}(x^*)\mathrm{d}W_l, \tag{3.18}$$

at $V \to \infty$. This is an Ornstein-Uhlenbeck process, an example of a Gaussian process.

The equilibrium state (S^*, x_1^*, x_2^*) is not a fixed point, and is dependent on the initial value. However, the value of S^* is constant, and the point (x_1^*, x_2^*) is kept on the centre manifold.

We assume the surviving species in the strong selection are two quasi-neutral species sharing the same smallest S^*. Given the results from Chapter

2, the coefficients in the Ornstein-Uhlenbeck process Eq. (3.18) satisfy,

$$\partial F(x^*) = \begin{pmatrix} -D - \sum_{k=1,2} \frac{m_k a_k x_k^*}{y_k (a_k + S^*)^2} & -\frac{d_1}{y_1} & -\frac{d_2}{y_2} \\ \frac{m_1 a_1 x_1^*}{(a_1 + S^*)^2} & 0 & 0 \\ \frac{m_2 a_2 x_2^*}{(a_2 + S^*)^2} & 0 & 0 \end{pmatrix} \tag{3.19}$$

and

$$\sum l l^T \beta_l = \mathbf{G} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2d_1 x_1^* & 0 \\ 0 & 0 & 2d_2 x_2^* \end{pmatrix}. \tag{3.20}$$

Substituting Eqs. (3.19) and (3.20) into Eq. (3.18), we obtain the diffusion approximation with expression:

$$\mathrm{d}\hat{S}(t) = \Big\{ \Big(-D - \sum \frac{x_k^* \gamma_k}{y_k} \Big) (\hat{S} - S^*) - \sum \frac{d_k}{y_k} \Big(\hat{x}_k - x_k^* \Big) \Big\} \mathrm{d}t \tag{3.21}$$

$$\mathrm{d}\hat{x}_k(t) = \gamma_k x_k^* (\hat{S} - S^*) \mathrm{d}t + V^{-\frac{1}{2}} \sqrt{d_k x_k^*} \Big(\mathrm{d}W_{k,1} - \mathrm{d}W_{k,-1} \Big) \tag{3.22}$$

as $V \to \infty$, where $\gamma_k = \frac{(m_k - d_k)^2}{m_k a_k}$ is the derivative of the birth rate with respect to the resource concentration at the equilibrium state $\frac{\mathrm{d}\beta_{k,1}}{\mathrm{d}\hat{S}}|_{S^*}$ and defined as trade-off parameter in last chapter.

By the law of large numbers, the limit behaviour of this diffusion is a linearised flow $\mathrm{d}\bar{x}(t) = \partial F(x^*) \Big(\bar{x}(t) - x^*(t) \Big) \mathrm{d}t$. With the modifications from the random walk to this diffusion process, the tendency of the walk to move back towards the equilibrium state is shown. The results of the numerical calculation is plotted in Figure 3.2 to illustrate this.

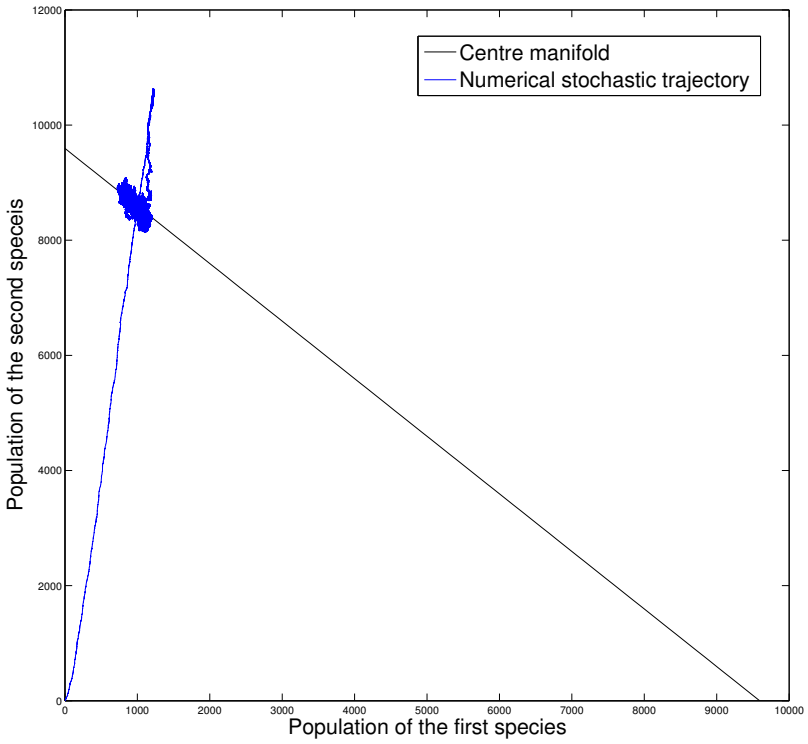

Figure 3.2: The trajectory of the stochastic process for the two quasi-neutral species around the centre manifold with initial value $N_1 = N_2 = 10$ and parameters $a_1 = 1 \cdot 10^{-6}$, $m_1 = 1$, $y_1 = 5 \cdot 10^9, a_2 = 2 \cdot 10^{-6}$, $m_1 = 1.925$, $y_1 = 5 \cdot 10^9$.

3.4.3 Weakness of the Diffusion Approximation

By the central limit theorem, the stochastic process converges to a Gaussian process with its linearised deterministic limit proving the characteristic of mean reverting. However, there exists a weakness in this diffusion approximation which will be unfolded in the following:

Firstly, the central limit theorem is applied to determine a Gaussian

process fluctuating round the deterministic limit which results from the law of large numbers. In this manner, our central limit theorem is conditioned over compact time intervals in the case of large but finite V. That is, asymptotically, the diffusion approximation and original stochastic process are essentially equivalent for compact time intervals, but not infinite time intervals. Consequently, a new diffusion approximation over an infinite time interval is need to be derived when we focusing on the long time dynamics.

Secondly, during the application of the central limit theorem, when the rescaled centred process

$$\chi^V(t) = \chi^V(0) + \sum_l lW_l^{(V)}\{\int_0^t \beta_l\Big(\hat{x}(s)\Big)\mathrm{d}s\} + \int_0^t \sqrt{V}\Big(F(\hat{x}(s)) - F(\bar{x}(s)\Big)\mathrm{d}s$$

converging to

$$X(t) = X(0) + \sum_l lW_l\{\int_0^t \beta_l(\bar{x}(s))\mathrm{d}s\} + \int_0^t \partial F(\bar{x}(s))X(s)\mathrm{d}s,$$

the last deterministic term on the right hand side is linearised, and the terms beyond the second order are dropped in the Taylor's expansion.

This results in a linearised process when we project the diffusion approximation onto the deterministic trajectory,

$$\begin{aligned}\mathrm{d}\bar{S}(t) &= \Big\{\Big(-D - \sum \frac{x_k^*\gamma_k}{y_k}\Big)(\bar{S} - S^*) - \sum \frac{d_k}{y_k}\Big(\bar{x}_k - x_k^*\Big)\Big\}\mathrm{d}t \\ \mathrm{d}\bar{x}_k(t) &= \gamma_k x_k^*(\bar{S} - S^*)\mathrm{d}t,\end{aligned}$$

which is actually the first order Taylor expansion of the deterministic ODE

system around the steady states.

The absence of the higher order terms will not result in a big influence on the dynamics over compact time intervals. However, when using this linearised deterministic limit to analysis the long time behaviour, the analytical results for the fixation probability do not match the numerical results very well, which will be shown in Chapter 4.

Above all, over compact time intervals, the Ornstein-Uhlenbeck process is sufficient to estimate the behaviour of the stochastic process. However, over a longer periods of time, the convergence from the stochastic dynamics to the deterministic dynamics will almost certainly not hold. In the following sections, we will derive a new diffusion approximation for $t \to \infty$ which will then correspond to a new deterministic trajectory along the centre manifold, and later determine how this stochastic population process will be trapped in to an absorption state. The method to derive the diffusion approximation is inspired by Dr. Todd Parsons.

3.5 Stochastic Process at a Long Time Scale and Its Projection Map onto the Deterministic Trajectory

In this section, we will by changing the time scale of the stochastic process to Vt, to find the stochastic process over an infinite time interval. The definition will be given in Section 3.5.1. To derive the diffusion approxima-

tion of the stochastic process at a long time scale, a mathematical trick will be used in Section 3.5.2. We will project the stochastic process onto the deterministic trajectory as $t \to \infty$, and prove that the projection map can well approximate the stochastic process. In Section 3.5.3, the derivatives of the projection map are calculated.

3.5.1 Stochastic Process at a Long Time Scale

Given infinite large V and hence large extinction time, we will focus on the long time behaviour. Define the Markov process on a longer time-scale with the following new definition:

$$\hat{s}(t) \stackrel{\text{def}}{=} \hat{S}(Vt), \tag{3.23}$$

$$\hat{w}(t) \stackrel{\text{def}}{=} \frac{1}{V}\hat{N}(Vt) = \hat{x}(Vt), \tag{3.24}$$

$$\hat{w}(0) = \bar{w}(0) = w,$$

$$\hat{s}(0) = s,$$

where the time scale is multiplied by the value of the volume which is close to infinity, and assume the initial state (s, w) stays on the centre manifold Ω.

Given the definition in Section 3.2, the long time scale process $\left(\hat{s}(t), \hat{w}(t)\right)$

is a solution of,

$$\begin{aligned}
\mathrm{d}\hat{s}(t) &= V\Big\{D(S^{\mathrm{in}}-\hat{s}) - \sum \frac{a_k\hat{s}}{y_k\Big(m_k+\hat{s}(t)\Big)}\hat{w}\Big\}\mathrm{d}t \\
\mathrm{d}\hat{w}_k(t) &= V\Big\{\Big(\frac{a_k\hat{s}}{m_k+\hat{s}(t)} - \mu_k - D\Big)\hat{w}_k\Big\}\mathrm{d}t \\
&\quad + \sum_l lV^{-1}\tilde{Y}_l\{V^2\beta_l\Big(\hat{S}(t),\hat{x}(t)\Big)\mathrm{d}t\}
\end{aligned} \tag{3.25}$$

with initial state (s,w), and $Y_l(s(\Delta t))$ as an independent standard Poisson process with $s(\Delta t)$ as its expected number of events during time Δt. Set $B_l^V(u) = V^{-\frac{1}{2}}\tilde{Y}_l(Vu)$, then $B_{k,l}^V(u) = \frac{1}{V}\tilde{Y}_l(V^2u) \Rightarrow B_{k,l}(u)$ are independent Brownian motions. Therefore, the stochastic part of Eq.(3.25) approximates to $\sqrt{b_k\hat{w}_k(t)}\mathrm{d}B_{k,1} - \sqrt{d_k\hat{w}_k(t)}\mathrm{d}B_{k,-1}$.

Over an infinite time interval, the law of large numbers and the central limit theorem are inapplicable, and the limit of the stochastic process is not the smooth solution of the ODE system at time t. Nevertheless, the diffusion approximations derived in the previous section may still be used to describe the dynamics over an appropriate small compact time interval $(t-\frac{1}{V}\Delta t,\ t+\frac{1}{V}\Delta t)$ in the infinite time interval,

$$\mathrm{d}\hat{s}(t) = V\Big\{\Big(-D-\sum\frac{w_k^*(t)\gamma_k}{y_k}\Big)(\hat{s}-s^*) - \sum\frac{d_k}{y_k}\Big(\hat{w}_k - w_k^*(t)\Big)\Big\}\mathrm{d}t \tag{3.26}$$

$$\mathrm{d}\hat{w}_k(t) = V\Big\{\gamma_k w_k^*(t)(\hat{s}-s^*)\Big\}\mathrm{d}t + \sqrt{d_k w_k^*(t)}(\mathrm{d}B_{k,1} - \mathrm{d}B_{k,-1}), \tag{3.27}$$

where the equilibrium state $(s^*, w_1^*(t), w_2^*(t))$ is no longer a fixed point dependent on the initial states, but a flow on the centre manifold. However, no

matter where the stochastic drift moves to, from Eqs. (3.26) and (3.27), we can see that at any time there is a strong attraction pulling the trajectories back to the centre manifold. This attraction force is V times the one at the normal time scale, and ensures that the stochastic process $\left(\hat{s}(t), \hat{w}(t)\right)$ converges to the centre manifold $\left(s^*, w^*(t)\right)$. Therefore, in the following section, in order to explain the stochastic dynamics over infinite time intervals, we will focus on determining the long time behaviour on the centre manifold.

3.5.2 Projection Map onto the Deterministic Trajectory

Define $\pi(s, w)$ as the value on the deterministic system as $t \to \infty$, with the initial state (s, w). To make the definition more precise, let

$$\pi_0(s, w) \stackrel{\text{def}}{=} \lim_{t\to\infty} \bar{S}(t, s, w),$$
$$\pi_k(s, w) \stackrel{\text{def}}{=} \lim_{t\to\infty} \bar{x}_k(t, s, w).$$

where $\left(\bar{S}(t, s, w), \bar{x}_k(t, s, w)\right)$ is the unique solution of the deterministic ODE system at the normal time scale (defined in Section 2.3) with initial conditions $\bar{S}(0) = s$ and $\bar{x}_k = w_k$. With this definition, π depends only on the initial condition and not on time. If the initial value is on the centre manifold (S^*, x^*), then it will remain there indefinitely, namely $\pi(S^*, x^*) = (S^*, x^*)$.

Projection Map

Substitute the initial condition with the stochastic process at a long time scale, then $\pi\Big(\hat{s}(t), \hat{w}(t)\Big) = \Big(\pi_0\big(\hat{s}(t), \hat{w}(t)\big), \pi_1\big(\hat{s}(t), \hat{w}(t)\big), \pi_2\big(\hat{s}(t), \hat{w}(t)\big)\Big)$ is the projection map of $\Big(\hat{s}(t), \hat{w}(t)\Big)$ onto the deterministic trajectory as $t \to \infty$.

As was explained in the previous subsection, due to an extremely strong attraction towards the centre manifold, the stochastic process at a long time scale will always move immediately back to the centre manifold. Therefore

$$\Big(\hat{s}(t),\ \hat{w}(t)\Big) \Rightarrow \Big(S^*,\ w^*(t)\Big) \quad \text{as} \quad V \to \infty, \tag{3.28}$$

and

$$\pi_k\Big(\hat{s}(t), \hat{w}(t)\Big) \Rightarrow \pi_k\Big(S^*, w^*(t)\Big), \tag{3.29}$$

where $\Big(S^*, w^*(t)\Big)$ is the trajectory on the centre manifold. As the nature of our deterministic dynamics, if the process start from an equilibrium state, it will remain there, then

$$\pi\Big(S^*, w^*(t)\Big) = \Big(S^*,\ w^*(t)\Big). \tag{3.30}$$

Given Eqs.(3.28),(3.29) and (3.30), we yield,

$$\pi\Big(\hat{s}(t), \hat{w}(t)\Big) - \Big(\hat{s}(t),\ \hat{w}(t)\Big) \Rightarrow 0. \tag{3.31}$$

Hence, this projection map onto the deterministic trajectory as $t \to \infty$ is sufficient to predict the behaviour of the stochastic process at the long time

scale. If the explicit expression of $\pi\Big(\hat{s}(t), \hat{w}(t)\Big)$ could be derived, then it will be the diffusion approximation of the stochastic process at a long time scale.

In the following Section 3.6, with the help of Itô's transformation and Thomas Kurtz's weak convergence, $\pi_k\Big(\hat{s}(t), \hat{w}(t)\Big)$ can be calculated, where $\frac{\mathrm{d}\pi(\hat{w})}{\mathrm{d}\hat{w}}$ is needed. Thus, before we derive the diffusion approximation, it is necessary to calculate the derivatives of the projection map.

3.5.3 Derivatives of the Projection Map

In this section, we will derive the first and second derivatives of the projection map with respect to the initial values $\frac{\mathrm{d}\pi(\hat{w})}{\mathrm{d}\hat{w}}$, which are needed in the later analysis to calculate the explicit expression of the diffusion approximation.

As the long time result of deterministic dynamics, $\pi(s, w)$ will stay indefinitely on the centre manifold, its value satisfies Eq. (2.9):

$$\pi_0 = S^*$$
$$(S^{\mathrm{in}} - S^*)D - \sum_k^2 \frac{\pi_k}{y_k} d_k = 0.$$

By taking derivatives of the above equation with respect to the initial state w_k, the relationships between the first derivatives may be expressed as:

$$\frac{\mathrm{d}\pi_1}{\mathrm{d}w_k}\frac{d_1}{y_1} + \frac{\mathrm{d}\pi_2}{\mathrm{d}w_k}\frac{d_2}{y_2} = 0, \tag{3.32}$$

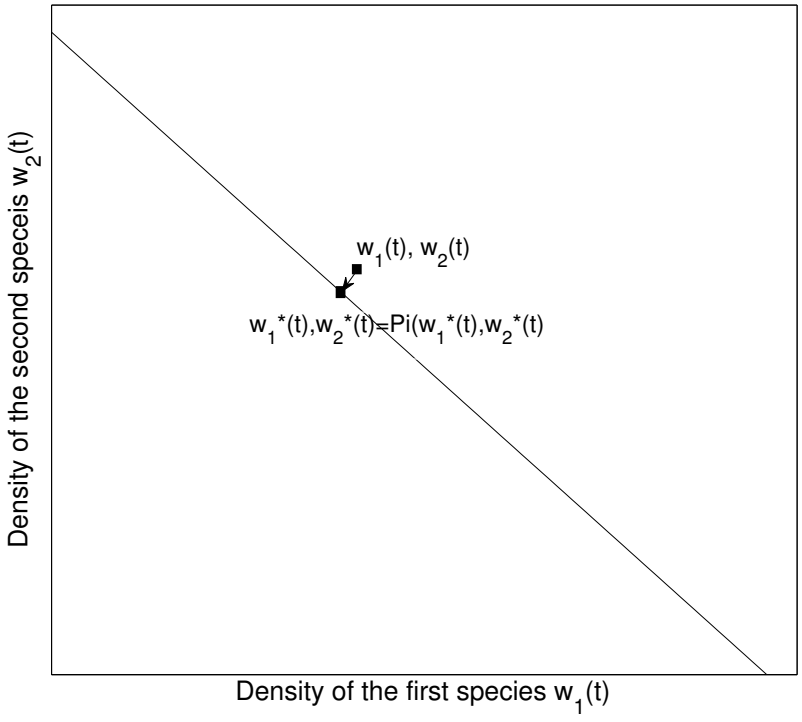

Figure 3.3: The point $(\hat{w}_1(t), \hat{w}_2(t))$ jumps to the centre manifold $(\hat{w}_1^*(t), \hat{w}_2^*(t))$ immediately. In deterministic trajectory, the long time results $\pi(\hat{w}_1^*(t), \hat{w}_2^*(t))$ remains on the initial point $(\hat{w}_1^*(t), \hat{w}_2^*(t))$. Therefore the trajectory of projection map $\pi(\hat{w}_1(t), \hat{w}_2(t))$ is sufficient to tell the behaviour of the process $(\hat{w}_1(t), \hat{w}_2(t))$.

which implies that,

$$\frac{\mathrm{d}\pi_1}{\mathrm{d}w_k} = -\frac{\kappa_2}{\kappa_1}\frac{\mathrm{d}\pi_2}{\mathrm{d}w_k} \tag{3.33}$$

$$\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_k^2} = -\frac{\kappa_2}{\kappa_1}\frac{\mathrm{d}^2\pi_2}{\mathrm{d}w_k^2}, \tag{3.34}$$

where $\kappa_k = \frac{d_k}{y_k}$.

In Section 3.4.2, when the central limit theorem is applied to derive the diffusion approximation, an approximation on the deterministic trajectory is obtained, which is actually the first order Taylor expansion of the deterministic ODE system around the equilibrium state. Then the expression of the deterministic trajectory $\Big(\bar{s}(t), \bar{x}(t)\Big)$ can linearly approximated as,

$$\mathrm{d}\bar{S}(t) = \Big\{\Big(- D - \sum \frac{x_k^*(t)\gamma_k}{y_k}\Big)\Big(\bar{S}(t) - S^*\Big) - \sum \frac{d_k}{y_k}\Big(\bar{x}_k - x_k^*(t)\Big)\Big\}\mathrm{d}t \tag{3.35}$$

$$\mathrm{d}\bar{x}_k(t) = \Big\{\gamma_k x_k^*(t)\Big(\bar{S}(t) - S^*\Big)\Big\}\mathrm{d}t \tag{3.36}$$

with the boundaries $\bar{x}_k(t=0) = w_k$ and $\bar{x}_k(t=\infty) = \pi_k$.

By applying Eq. (3.36)

$$\frac{\mathrm{d}\bar{x}_1}{\mathrm{d}\bar{x}_2} = \frac{\Big\{\gamma_1 x_1^*(t)\Big(\bar{S}(t) - S^*\Big)\Big\}\mathrm{d}t}{\Big\{\gamma_2 x_2^*(t)\Big(\bar{S}(t) - S^*\Big)\Big\}\mathrm{d}t} = \frac{\gamma_1 \bar{x}_1}{\gamma_2 \bar{x}_2},$$

with the boundaries, we obtain the relationship:

$$w_1^{\gamma_2}\pi_2^{\gamma_1} = w_2^{\gamma_1}\pi_1^{\gamma_2}. \tag{3.37}$$

Furthermore, we differentiate Eq. (3.37) with respect to w_1,

$$\gamma_2 w_1^{\gamma_2 - 1}\pi_2^{\gamma_1} + \gamma_1 w_1^{\gamma_2}\pi_2^{\gamma_1 - 1}\frac{\mathrm{d}\pi_2}{\mathrm{d}w_1} = \gamma_2 w_2^{\gamma_1}\pi_1^{\gamma_2 - 1}\frac{\mathrm{d}\pi_1}{\mathrm{d}w_1},$$

and substitute Eqs. (3.32) and (3.37) into this equation. Then the first

derivative $\frac{\mathrm{d}\pi_1}{\mathrm{d}w_1}$ is found,

$$\frac{\mathrm{d}\pi_1}{\mathrm{d}w_1} = \frac{\gamma_2\kappa_2\pi_1\pi_2}{(\gamma_1\kappa_1\pi_1 + \gamma_2\kappa_2\pi_2)w_1}.$$

Substituting the initial value with the stochastic process at a long time scale $\hat{w}$, with $\hat{w}_k \Rightarrow \pi_k(\hat{w})$ from Eq.(3.31), we have the final results for the derivatives of the projection map,

$$\frac{\mathrm{d}\pi_1(\hat{w})}{\mathrm{d}\hat{w}_1} \Rightarrow \frac{\gamma_2\kappa_2\pi_2}{\gamma_1\kappa_1\pi_1 + \gamma_2\kappa_2\pi_2}. \tag{3.38}$$

In a similar way,

$$\frac{\mathrm{d}\pi_1(\hat{w})}{\mathrm{d}\hat{w}_2} \Rightarrow -\frac{\gamma_1\kappa_2\pi_1}{\gamma_1\kappa_1\pi_1 + \gamma_2\kappa_2\pi_2}. \tag{3.39}$$

Differentiating Eq. (3.37) again, we can obtain the second derivatives,

$$\begin{aligned}
&\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_1^2} \Rightarrow \frac{-\kappa_1\kappa_2\gamma_1\gamma_2\pi_2}{(\kappa_1\pi_1\gamma_1 + \kappa_2\pi_2\gamma_2)^2}(1 + \frac{\gamma_2(\kappa_1\pi_1 + \kappa_2\pi_2)}{\kappa_1\pi_1\gamma_1 + \kappa_2\pi_2\gamma_2}),\\
&\frac{\mathrm{d}^2\pi_2}{\mathrm{d}w_1^2} = -\frac{\kappa_1}{\kappa_2}\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_1^2},\\
&\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_2^2} \Rightarrow \frac{\kappa_2^2\gamma_1\gamma_2\pi_1}{(\kappa_1\pi_1\gamma_1 + \kappa_2\pi_2\gamma_2)^2}(1 + \frac{\gamma_1(\kappa_1\pi_1 + \kappa_2\pi_2)}{\kappa_1\pi_1\gamma_1 + \kappa_2\pi_2\gamma_2},\\
&\frac{\mathrm{d}^2\pi_2}{\mathrm{d}w_2^2} = -\frac{\kappa_1}{\kappa_2}\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_2^2}.
\end{aligned}$$

Using these results, the diffusion approximation at a long time scale and the fixation probability will be calculated in the later sections. However, when comparing the analytical fixation probabilities with the simulation results in Chapter 4, we found large discrepancies. These difference did not

disappear as the population was increased, suggesting that a more accurate approximation of the derivatives is necessary.

3.5.3.1 Corrections to the Derivatives of the Projection Map

In Eqs. (3.35) and (3.36), we use the first-order Taylor expansion around the equilibrium state to express the behaviour on the deterministic trajectory. The ignorance of the higher order terms may result in inaccuracy of the final analytical prediction.

Thus in this subsection, we will use the second-order Taylor expansion of the deterministic ODE system around the equilibrium state, to calculate the derivatives of the projection map. The results calculated in this subsection will be applied to correct the fixation probabilities which will be demonstrated to match the numerical results better in Chapter 4.

The second order Taylor series expansion of our deterministic dynamics can be written as

$$\frac{\mathrm{d}\bar{x}(t)}{\mathrm{d}t} = \partial F(x^*)\Big(\bar{x}(t) - x^*\Big) + \sum_{i,j} \partial_i \partial_j F(x^*)\Big(\bar{x}_i(t) - x_i^*\Big)\Big(\bar{x}_j(t) - x_j^*\Big).$$

Expanding it in detail, we have

$$\frac{\mathrm{d}\bar{S}}{\mathrm{d}t} = (-D - \sum \frac{\gamma_k \bar{x}_k}{y_k})(\bar{S} - S^*) + \sum \frac{d_k \bar{x}_k}{m_k y_k S^*}(\bar{S} - S^*)^2 \tag{3.40}$$

$$\frac{\mathrm{d}\bar{x}_k}{\mathrm{d}t} = \gamma_k \bar{x}_k (\bar{S} - S^*) - \frac{d_k}{m_k S^*}\gamma_k \bar{x}_k (\bar{S} - S^*)^2 \tag{3.41}$$

with boundaries at $(\bar{S}, \bar{x}_k)(t = 0) = (\hat{s}, \hat{w}_k)$, $(\bar{S}, \bar{x}_k)(t = \infty) = (\pi_0, \pi_k)$.

Cautions in the calculation Since the initial condition $\bar{S}(t=0) = \hat{s} \Rightarrow S^*$, the trajectory of $\bar{S}(t)$ is kept constant all through. Therefore, in the following calculation, the integral of functions of $\bar{S}(t) - S^*$ from $t = 0$ to ∞ can always be vanished, i.e

$$\int_0^\infty f\Big(\bar{S}(t) - S^*\Big)\mathrm{d}t = 0. \tag{3.42}$$

First Derivatives

Taking the first derivative of Eq. (3.41) with respect to w_1, we have

$$\begin{aligned}
\frac{1}{\gamma_1}\frac{\mathrm{d}^2 \ln \bar{x}_1}{\mathrm{d}t\mathrm{d}w_1} &= \frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1} - \frac{2d_1(\bar{S} - S^*)}{m_1 S^*}\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1} \\
\frac{1}{\gamma_2}\frac{\mathrm{d}^2 \ln \bar{x}_2}{\mathrm{d}t\mathrm{d}w_1} &= \frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1} - \frac{2d_2(\bar{S} - S^*)}{m_2 S^*}\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1},
\end{aligned}$$

with its integral system from $t = 0$ to $t = \infty$:

$$\frac{1}{\gamma_1\pi_1}\frac{\mathrm{d}\pi_1}{\mathrm{d}w_1} - \frac{1}{\gamma_1 w_1} = \int_0^\infty \frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1}\mathrm{d}t - \int_0^\infty \frac{2d_1(\bar{S} - S^*)}{m_1 S^*}\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1}\mathrm{d}t \tag{3.43}$$

$$\frac{1}{\gamma_2\pi_2}\frac{\mathrm{d}\pi_2}{\mathrm{d}w_1} = \int_0^\infty \frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1}\mathrm{d}t - \int_0^\infty \frac{2d_2(\bar{S} - S^*)}{m_2 S^*}\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1}\mathrm{d}t. \tag{3.44}$$

Given Eq. (3.42), the second terms on the right hand side of Eqs. (3.43) and (3.44), $\int_0^\infty \frac{2d_k(\bar{S}(t) - S^*)}{m_k S^*}\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1}\mathrm{d}t$, are close to 0, thus the second order term of the Taylor expansion in Eq. (3.41) may be neglected.

Rearranging Eq.(3.43) and (3.44), we have

$$\frac{1}{\gamma_1\pi_1}\frac{\mathrm{d}\pi_1}{\mathrm{d}w_1} - \frac{1}{\gamma_1 w_1} = \frac{1}{\gamma_2\pi_2}\frac{\mathrm{d}\pi_2}{\mathrm{d}w_1}.$$

With the relationship on the centre manifold in Eq. (3.33) $\frac{\mathrm{d}\pi_1}{\mathrm{d}w_k} = -\frac{\kappa_2}{\kappa_1}\frac{\mathrm{d}\pi_2}{\mathrm{d}w_k}$, the results for the first derivatives of the projection map will not differ from the results derived by the first order Taylor expansion, which is shown in Eqs. (3.38) and (3.39).

Second Derivatives

The first derivatives of the projection map are obtained from solving the Eqs. (3.43) and (3.44) with $\int_0^\infty \frac{2d_k(\bar{S}(t)-S^*)}{m_k S^*}\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1}\mathrm{d}t = 0$. To derive the second derivatives of the projection map, we take derivatives of the system. The non-zero value of the first derivative of $\int_0^\infty \frac{2d_k(\bar{S}(t)-S^*)}{m_k S^*}\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1}\mathrm{d}t$ in the system cause the second order term of the Taylor expansion in Eq. (3.41) to be no longer negligible. Therefore, the second derivatives of the projection map are not the same as in the last subsection.

The followings are the steps to calculate the second derivative $\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_1^2}$:

Firstly, differentiate Eq. (3.43) with respect to w_1:

$$\frac{1}{\gamma_1\pi_1}\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_1^2} - \frac{1}{\gamma_1\pi_1^2}(\frac{\mathrm{d}\pi_1}{\mathrm{d}w_1})^2 + \frac{1}{\gamma_1 w_1^2} = \int_0^\infty \frac{\mathrm{d}^2\bar{S}}{\mathrm{d}w_1^2}\mathrm{d}t - \int_0^\infty \frac{2d_1}{m_1 S^*}(\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1})^2\mathrm{d}t - \underbrace{\int_0^\infty \frac{2d_1(\bar{S}-S^*)}{m_1 S^*}\frac{\mathrm{d}^2\bar{S}}{\mathrm{d}w_1^2}\mathrm{d}t}_{\to 0}.$$

The third term on the right hand side can be omitted as previously, but not the second term. Similarly, the differentiation of Eq. (3.44) with respect to w_1 is:

$$\frac{1}{\gamma_2\pi_2}\frac{\mathrm{d}^2\pi_2}{\mathrm{d}w_1^2} - \frac{1}{\gamma_2\pi_2^2}(\frac{\mathrm{d}\pi_2}{\mathrm{d}w_1})^2 = \int_0^\infty \frac{\mathrm{d}^2\bar{S}}{\mathrm{d}w_1^2}\mathrm{d}t - \int_0^\infty \frac{2d_2}{m_2 S^*}(\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1})^2\mathrm{d}t.$$

After rearranging the above two equations, we have the following relationship for the second derivatives of the projection map:

$$\begin{aligned}\int_0^\infty \frac{\mathrm{d}^2\bar{S}}{\mathrm{d}w_1^2}\mathrm{d}t &= \frac{1}{\gamma_1\pi_1}\Big\{\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_1^2} + \gamma_1\pi_1\int_0^\infty \frac{2d_1}{m_1S^*}(\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1})^2\mathrm{d}t - \frac{1}{\pi_1}(\frac{\mathrm{d}\pi_1}{\mathrm{d}w_1})^2 + \frac{1}{\pi_1}\Big\} \\ &= \frac{1}{\gamma_2\pi_2}\Big\{\frac{\mathrm{d}^2\pi_2}{\mathrm{d}w_1^2} + \gamma_2\pi_2\int_0^\infty \frac{2d_2}{m_2S^*}(\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1})^2\mathrm{d}t - \frac{1}{\pi_2}(\frac{\mathrm{d}\pi_2}{\mathrm{d}w_1})^2\Big\}.\end{aligned}$$

Given the relationship in Eqs. (3.33) and (3.34) and the results for the first derivatives in Eqs.(3.38) and (3.39), the system of the second derivatives is yielded:

$$\begin{aligned}&\frac{\gamma_2\pi_2}{\gamma_1\pi_1}\Big\{\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_1^2} + \gamma_1\pi_1\int_0^\infty \frac{2d_1}{m_1S^*}(\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1})^2\mathrm{d}t - \frac{1}{\pi_1}(\frac{\mathrm{d}\pi_1}{\mathrm{d}w_1})^2 + \frac{1}{\pi_1}\Big\} \\ &\quad = \frac{\mathrm{d}^2\pi_2}{\mathrm{d}w_1^2} + \gamma_2\pi_2\int_0^\infty \frac{2d_2}{m_2S^*}(\frac{\mathrm{d}\bar{S}}{\mathrm{d}w_1})^2\mathrm{d}t - \frac{1}{\pi_2}(\frac{\mathrm{d}\pi_2}{\mathrm{d}w_1})^2 \\ &\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_1^2} = -\frac{\kappa_2}{\kappa_1}\frac{\mathrm{d}^2\pi_2}{\mathrm{d}w_1^2} \\ &\frac{\mathrm{d}\pi_1}{\mathrm{d}\hat{w}_1} = \frac{\gamma_2\kappa_2\pi_2}{\gamma_1\kappa_1\pi_1 + \gamma_2\kappa_2\pi_2} = -\frac{\kappa_2}{\kappa_1}\frac{\mathrm{d}\pi_2}{\mathrm{d}\hat{w}_1},\end{aligned}$$

with solution

$$\begin{aligned}\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_1^2} =& -\frac{\kappa_1\kappa_2\gamma_1\gamma_2\pi_2}{(\kappa_1\pi_1\gamma_1 + \kappa_2\pi_2\gamma_2)^2}(1 + \frac{\gamma_2(\kappa_1\pi_1 + \kappa_2\pi_2)}{\kappa_1\pi_1\gamma_1 + \kappa_2\pi_2\gamma_2}) \\ &+ \frac{2\kappa_2\gamma_1\gamma_2\pi_1\pi_2 D}{S^*(\sum \kappa_k\gamma_k\pi_k)}(\frac{1}{m_2} - \frac{1}{m_1})\int_0^\infty (\frac{\mathrm{d}\pi_0}{\mathrm{d}w_1})^2\mathrm{d}t. \qquad (3.45)\end{aligned}$$

Secondly, we will calculate the term $\int_0^\infty (\frac{\mathrm{d}\pi_0}{\mathrm{d}w_1})^2 \mathrm{d}t$ in Eq. (3.45):

$$\begin{aligned}
\int_0^\infty (\frac{\mathrm{d}\pi_0}{\mathrm{d}w_1})^2 \mathrm{d}t &= \int_0^\infty \frac{\mathrm{d}\pi_0}{\mathrm{d}w_1} \mathrm{d}\Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \\
&= \frac{\mathrm{d}\pi_0}{\mathrm{d}w_1} \int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big|_0^\infty - \int_0^\infty \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \mathrm{d} \frac{\mathrm{d}\pi_0}{\mathrm{d}w_1} \\
&= 0 - \int_0^\infty \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \frac{\mathrm{d}^2\pi_0}{\mathrm{d}w_1 \mathrm{d}t} \mathrm{d}t \\
&= \int_0^\infty \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \Big((D + \sum \frac{\gamma_k \pi_k}{y_k}) \frac{\mathrm{d}\pi_0}{\mathrm{d}w_1} + \underbrace{\sum \frac{\gamma_k}{y_k} \frac{\mathrm{d}\pi_k}{\mathrm{d}w_1} (\pi_0 - S^*)}_{\to 0} \Big) \mathrm{d}t \\
&= \int_0^\infty \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) D \frac{\mathrm{d}\pi_0}{\mathrm{d}w_1} \mathrm{d}t + \int_0^\infty \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \sum \frac{\gamma_k \pi_k}{y_k} \frac{\mathrm{d}\pi_0}{\mathrm{d}w_1} \mathrm{d}t + 0 \\
&= D \int_0^\infty \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \mathrm{d} \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \\
&\quad + \int_0^\infty \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \sum \frac{1}{y_k} \frac{\mathrm{d}^2\pi_k}{\mathrm{d}t \mathrm{d}w_1} \mathrm{d}t \\
&= \frac{D}{2} \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big)^2 \Big|_0^\infty + \int_0^\infty \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \mathrm{d} \Big(\sum \frac{1}{y_k} \frac{\mathrm{d}\pi_k}{\mathrm{d}w_1} \Big).
\end{aligned} \tag{3.46}$$

Given $\sum \frac{1}{y_k} \frac{\mathrm{d}\pi_k}{\mathrm{d}w_1} = 0$ in Eq. (3.32), the second term on the right hand side of Eq. (3.46) is 0, as

$$\begin{aligned}
\int_0^\infty \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \mathrm{d} \Big(\sum \frac{1}{y_k} \frac{\mathrm{d}\pi_k}{\mathrm{d}w_1} \Big) &= \Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1} \mathrm{d}u \Big) \Big(\sum \frac{1}{y_k} \frac{\mathrm{d}\pi_k}{\mathrm{d}w_1} \Big) \Big|_0^\infty \\
&\quad - \int_0^\infty \Big(\sum \frac{1}{y_k} \frac{\mathrm{d}\pi_k}{\mathrm{d}w_1} \Big) \Big(\frac{\mathrm{d}\pi_0(t)}{\mathrm{d}w_1} \Big) \mathrm{d}t \\
&= 0.
\end{aligned}$$

Using the Eq. (3.44), we yield the first term on the right hand side of Eq.

(3.46) as,

$$\frac{D}{2}\Big(\int_0^t \frac{\mathrm{d}\pi_0(u)}{\mathrm{d}w_1}\mathrm{d}u\Big)^2\Big|_0^\infty = \frac{D}{2}(\frac{1}{\gamma_2\pi_2}\frac{\mathrm{d}\pi_2}{\mathrm{d}w_1})^2 = \frac{D\kappa_1^2}{2(\sum\kappa_k\gamma_k\pi_k)^2}.$$

Therefore, the calculation of Eq.(3.46) is completed, and we obtain

$$\int_0^\infty (\frac{\mathrm{d}\pi_0}{\mathrm{d}w_1})^2\mathrm{d}t = \frac{D\kappa_1^2}{2(\sum\kappa_k\gamma_k\pi_k)^2}, \tag{3.47}$$

where $\kappa_k = \frac{d_k}{y_k}$.

Lastly, by substituing Eq. (3.47) into the Eq. (3.45), the corrected results for $\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_1^2}$ is obtained,

$$\begin{aligned}\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_1^2} =& -\frac{\kappa_1\kappa_2\gamma_1\gamma_2\pi_2}{(\kappa_1\pi_1\gamma_1+\kappa_2\pi_2\gamma_2)^2}\Big(1+\frac{\gamma_2(\kappa_1\pi_1+\kappa_2\pi_2)}{\kappa_1\pi_1\gamma_1+\kappa_2\pi_2\gamma_2}\Big)\\ &+\gamma_1\pi_1\frac{D^2}{S^*}\frac{\kappa_1^2\kappa_2\gamma_2\pi_2}{(\sum\kappa_k\gamma_k\pi_k)^3}(\frac{1}{m_2}-\frac{1}{m_1}).\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\mathrm{d}^2\pi_1}{\mathrm{d}w_2^2} =& \frac{\kappa_2^2\gamma_1\gamma_2\pi_1}{(\kappa_1\pi_1\gamma_1+\kappa_2\pi_2\gamma_2)^2}\Big(1+\frac{\gamma_1(\kappa_1\pi_1+\kappa_2\pi_2)}{\kappa_1\pi_1\gamma_1+\kappa_2\pi_2\gamma_2}\Big)\\ &+\gamma_1\pi_1\frac{D^2}{S^*}\frac{\kappa_2^3\gamma_2\pi_2}{(\sum\kappa_k\gamma_k\pi_k)^3}(\frac{1}{m_2}-\frac{1}{m_1}).\end{aligned}$$

3.6 Diffusion Approximation at a Long Time Scale

In the last section, it is explained that the projection map can well approximate the stochastic process at a long time scale. Therefore, in this section, we will derive the diffusion approximation of the projection map,

which is the diffusion approximation of the stochastic process at a long time scale.

3.6.1 Diffusion Approximation of the Species Densities

With the help of Itô's transformation[2], the Ito stochastic integral equation of the projection map $\pi_k\Big(\hat{w}(t)\Big)$ can be generated immediately, which is

$$\pi_k\Big(\hat{w}(t)\Big) = \pi_k\Big(\hat{w}(0)\Big) + \sum_i^2 \int_0^t (\partial_i \pi_k)\Big(\hat{w}(s)\Big) \mathrm{d}\hat{w}_i(s) + \frac{1}{2}\sum_{i,j} \int_0^t (\partial_i \partial_j \pi_k)\Big(\hat{w}(s)\Big) \mathrm{d}[\hat{w}_i, \hat{w}_j]_s + \epsilon(t), \quad (3.48)$$

with the quadratic covariation[3]

$$[\hat{w}_i, \hat{w}_j]_s = [M_i, M_j]_s.$$

M_k is the stochastic term of $\hat{w}_k$ in Eq. (3.27),

$$M_k = \sqrt{d_k w_k^*(t)}\Big(\mathrm{d}B_{k,1} - \mathrm{d}B_{k,-1}\Big),$$

satisfying the quadratic covariation $[M_k, M_j]_{i\neq j} = 0$, and the quadratic variation $[M_i, M_i] = [M_i] = 2d_i w_i^*$.

[2]Itô's transformation is an identity used in Itô's calculus to find the differential of a time-dependent function of a stochastic process. It is memorised by forming the Taylor series expansion of the function up to its second derivatives and identifying the square of an increment in the Wiener process with an increment in time. (from Wiki)

[3]The quadratic variation is $[X, Y]_t = \lim_{i\to\infty} \sum_{t_i<t} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i})$

Expanding Eq. (3.48) in detail, gives

$$\pi_k\Big(\hat{w}(t)\Big) = \pi_k\Big(\hat{w}(0)\Big) + \underbrace{\int_0^t (D\pi_k)\Big(\hat{w}(s)\Big)\cdot F\Big(\hat{w}(s)\Big)\mathrm{d}s}_{\text{Finite Variation Component}} + \underbrace{\sum_i^2 \int_0^t (\partial_i \pi_k)\Big(\hat{w}(s)\Big)\mathrm{d}M_i(s)}_{\text{Martingale Component}}$$
$$+ \underbrace{\frac{1}{2}\sum_i^2 \int_0^t (\partial_i^2 \pi_k)\Big(\hat{w}(s)\Big)\mathrm{d}[M_i]_s}_{\text{Quadratic Variation Component}} + \epsilon(t). \tag{3.49}$$

In order to derive the approximation of Eq. (3.49), the weak convergence for the stochastic integrals and stochastic differential equations from Kurtz and Protter's paper [37] will be applied. This states that if (U_n, X_n, Y_n) satisfies $X_n(t) = U_n(t) + \int_0^t F_n(X_n, s-)dY_n(s)$, and $(U_n, Y_n, F_n) \Rightarrow (U, Y, F)$, then X_n converges to a solution X of the obvious limiting stochastic differential equation $X(t) = U(t) + \int_0^t F(X, s-)dY(s)$.

To apply this weak convergence, firstly, each component of the Eq. (3.49) will be approximated one by one in the following:

Finite Variation Component

$$\begin{aligned}&\int_0^t (D\pi_k)\Big(\hat{w}(s)\Big)\cdot F\Big(\hat{w}(s)\Big)\mathrm{d}s \\ &= \int_0^t \frac{\mathrm{d}\pi_k}{\mathrm{d}\hat{w}_1}F_1(\hat{w}) + \frac{\mathrm{d}\pi_k}{\mathrm{d}\hat{w}_2}F_2(\hat{w})\mathrm{d}s\end{aligned}$$

Using the results of the first derivatives of the projection map in Eqs. (3.38) and (3.39), we have

$$\frac{\mathrm{d}\pi_k}{\mathrm{d}\hat{w}_1}F_1(\hat{w}) + \frac{\mathrm{d}\pi_k}{\mathrm{d}\hat{w}_2}F_2(\hat{w}) \Rightarrow 0.$$

Thus the finite variation component of $\pi_k\Big(\hat{w}(t)\Big)$ vanishes, as $V \to \infty$, and it means that the deterministic reaction function F_k will have no influence on the final stochastic diffusion approximation in this model.

Quadratic Variation Component

$$\frac{1}{2}\sum_{i}^{2}\int_0^t (\mathrm{d}_i^2\pi_k)\Big(\hat{w}(s)\Big)\mathrm{d}[M_i]_s$$

As a *càdlàg* finite variation process, $\hat{w}_i$ has its quadratic variation equal to the sum of the square of the jumps of $\hat{w}_i$. This gives the result that $[M_i]_t \Rightarrow \int_0^t 2d_k w_k^*(s)ds$. Thus when $V \to \infty$, the convergence of the quadratic variation component is obtained as,

$$\frac{1}{2}\sum_{i}^{2}\int_0^t (\mathrm{d}_i^2\pi_k)\Big(\hat{w}(s)\Big)\mathrm{d}[M_i]_s \Rightarrow \int_0^t (\mathrm{d}_1^2\pi_k)d_1 w_1^*(s) + (\mathrm{d}_2^2\pi_k)d_2 w_2^*(s)\ \mathrm{d}s$$

with death rate $d_k = D + \mu_k$.

Martingale Component

$$\sum_{i}^{2}\int_0^t (\mathrm{d}_i\pi_k)\Big(\hat{w}(s)\Big)\mathrm{d}M_i(s)$$

As above, while $V \to \infty$

$$\begin{aligned}\sum_{i}^{2}\int_0^t (\mathrm{d}_i\pi_k)\Big(\hat{w}(s)\Big)\mathrm{d}M_i(s) \Rightarrow \sum_{l}\int_0^t (\mathrm{d}_1\pi_k)l\Big(d_1 w_1^*(s)\Big)^{\frac{1}{2}}\mathrm{d}W_{1,l}(s) \\ + (\mathrm{d}_2\pi_k)l\Big(d_2 w_2^*(s)\Big)^{\frac{1}{2}}\mathrm{d}W_{2,l}(s)\end{aligned}$$

Remainder terms $|\epsilon(t)|$ weakly converges to 0 as $V \to \infty$ [51].

As above, the only contribution from the deterministic reaction, i.e., the finite variation component, vanishes in the weak convergence. This leaves the martingale component and the quadratic variation component in the diffusion, both resulting from stochastic drift. Therefore, in our model with large but finite population, although the level of stochastic drift is tiny, it is an important mechanism and has significant influence over an infinite time interval, causing the departure of the process from the deterministic dynamics.

Finally, by applying the weak convergence in [37], $\pi(\hat{w})$ converges weakly to the diffusion process $\Pi(w,t) = \Big(\Pi_1(w,t), \Pi_2(w,t)\Big)$, which can be expressed in SDEs as,

$$\begin{aligned}
d\Pi_1 &= \Big\{(\mathrm{d}_1^2\Pi_1)d_1\Pi_1 + (\mathrm{d}_2^2\Pi_1)d_2\Pi_2\Big\}dt \\
&\quad + (\mathrm{d}_1\Pi_1)\sqrt{d_1\Pi_1}\Big(dW_{1,1} - dW_{1,-1}\Big) + (\mathrm{d}_2\Pi_1)\sqrt{d_2\Pi_2}\Big(dW_{2,1} - dW_{2,-1}\Big) \\
&= \frac{\gamma_1\gamma_2\kappa_2\Pi_1\Pi_2}{(\gamma_1\kappa_1\Pi_1 + \gamma_2\kappa_2\Pi_2)^2}\Big[(\kappa_2 d_2 - \kappa_1 d_1) + \frac{(\kappa_1\Pi_1 + \kappa_2\Pi_2)(\gamma_1\kappa_2 d_2 - \gamma_2\kappa_1 d_1)}{\gamma_1\kappa_1\Pi_1 + \gamma_2\kappa_2\Pi_2}\Big]dt \\
&\quad + \frac{\gamma_2\kappa_2\Pi_2\sqrt{d_1\Pi_1}}{\gamma_1\kappa_1\Pi_1 + \gamma_2\kappa_2\Pi_2}\Big(dW_{1,1} - dW_{1,-1}\Big) - \frac{\gamma_1\kappa_2\Pi_1\sqrt{d_2\Pi_2}}{\gamma_1\kappa_1\Pi_1 + \gamma_2\kappa_2\Pi_2}\Big(dW_{2,1} - dW_{2,-1}\Big)
\end{aligned} \tag{3.50}$$

$$\begin{aligned}
d\Pi_2 &= \Big\{(\mathrm{d}_1^2\Pi_2)d_1\Pi_1 + (\mathrm{d}_2^2\Pi_2)d_2\Pi_2\Big\}dt \\
&\quad + (\mathrm{d}_1\Pi_2)\sqrt{d_1\Pi_1}\Big(dW_{1,1} - dW_{1,-1}\Big) + (\mathrm{d}_2\Pi_2)\sqrt{d_2\Pi_2}\Big(dW_{2,1} - dW_{2,-1}\Big), \\
&= -\frac{\kappa_1}{\kappa_2}d\Pi_1
\end{aligned}$$

where $d_k = b_k^* = \frac{m_k s^*}{a_k + s^*}$, and the initial state $(s, w) \in \Omega$.

The derivatives of the projection map were calculated in Subsection 3.5.3, where two different sets are stated. The first set of the derivatives derived by the linearised expression of the deterministic system are generally used in Eq. (3.50) to yield the final expression for the diffusion approximation over a infinite time interval. In the later discussion, the diffusion approximation will be used to yield the fixation probabilities. Then, special care needs to be taken when comparing the analytical results with the numerical results of the fixation probabilities. If differences are observed, the corrected set of the derivatives of the projection map (in Subsection 3.5.3.2) have to be applied to ensure the analytical prediction is accurate.

3.6.2 Diffusion Approximation of the Relative Abundance of the First Species

As our model has only two quasi-neutral species, analysing the relative abundance of the first species is sufficient to understand the behaviour of the whole population. Given the diffusion approximation of the species densities, the relative abundance of first species at the long time scale is

$$P(t) = \frac{\Pi_1(t)}{\Pi_1(t) + \Pi_2(t)}.$$

In this subsection, we will calculate the diffusion approximation of $P(t)$.

Since $\Big(\Pi_1(t), \Pi_2(t)\Big)$ is kept on the centre manifold,

$$0 = (S^{\text{in}} - S^*)D - (\kappa_1 \Pi_1 + \kappa_2 \Pi_2),$$

then $\Pi_2 = \frac{(S^{\text{in}}-S^*)D-\kappa_1\Pi_1}{\kappa_2}$, and

$$P(t) = \frac{\kappa_2\Pi_1(t)}{(S^{\text{in}} - S^*)D + (\kappa_2 - \kappa_1)\Pi_1(t)} \tag{3.51}$$

with $\kappa_k = \frac{d_k}{y_k}$.

Applying Itô's formula on Eq. (3.51), we get,

$$dP = \frac{\mathrm{d}P}{\mathrm{d}\Pi_1}(\mathrm{d}\Pi_1) + \frac{1}{2}\frac{\mathrm{d}^2P}{\mathrm{d}\Pi_1^2}(\mathrm{d}\Pi_1)^2, \tag{3.52}$$

with the derivatives

$$\frac{\mathrm{d}P}{\mathrm{d}\Pi_1} = \frac{\kappa_2(S^{\text{in}} - S^*)D}{\Big((S^{\text{in}} - S^*)D + (\kappa_2 - \kappa_1)\Pi_1(t)\Big)^2} = \frac{\kappa_2(1-P) + \kappa_1 P}{\kappa_2(\Pi_1 + \Pi_2)}$$
$$\frac{\mathrm{d}^2P}{\mathrm{d}\Pi_1^2} = 2\frac{(\kappa_2(1-P) + \kappa_1 P)(\kappa_1 - \kappa_2)}{\kappa_2^2(\Pi_1 + \Pi_2)^2},$$

where the total population

$$\Theta = \Pi_1 + \Pi_2 = \frac{(S^{\text{in}} - S^*)D}{\kappa_1 P + \kappa_2(1-P)}.$$

Denote the initial relative abundance as

$$P(0) = p = \frac{w_1}{w_1 + w_2}.$$

In order to completely expand the Eq. (3.52), it is necessary to use the limit that : for any independent standard Bronian process B_k, as $\mathrm{d}t$ tends to 0, the terms of order $\mathrm{d}t^2$, $\mathrm{d}B_1\mathrm{d}B_2$ and $\mathrm{d}t\mathrm{d}B_k$ will disappear but the $\mathrm{d}B_k^2$ term tends to $\mathrm{d}t$.

Therefore, given dΠ in the last subsection, the diffusion approximation of the relative abundance of the first species in Eq. (3.52) can be rewritten as a SDE:

$$dP = b(P)dt + a(P)dB_t,$$

where the drift coefficient $b(P)$ is

$$\begin{aligned} b(P) &= \frac{\mathrm{d}P}{\mathrm{d}\Pi_1}\Big\{d_1\pi_1\mathrm{d}_1^2\pi_1 + d_2\pi_2\mathrm{d}_2^2\pi_1 + \frac{2(\kappa_1-\kappa_2)}{\kappa_2(\Pi_1+\Pi_2)}\Big(d_1\pi_1(\mathrm{d}_1\pi_1)^2 + d_2\pi_2(\mathrm{d}_2\pi_1)^2\Big)\Big\} \\ &= \frac{\mathrm{d}P}{\mathrm{d}\Pi_1}\frac{P(1-P)\kappa_2}{\Big(\gamma_1\kappa_1P + \gamma_2\kappa_2(1-P)\Big)^2}\Big\{(\kappa_2d_2-\kappa_1d_1)\gamma_1\gamma_2 \\ &+ \gamma_1\gamma_2\frac{\Big(\kappa_1P+\kappa_2(1-P)\Big)(\gamma_1\kappa_2d_2-\gamma_2\kappa_1d_1)}{\kappa_1\gamma_1P+\kappa_2\gamma_2(1-P)} + 2(\kappa_1-\kappa_2)\Big(\gamma_2^2d_1(1-P)+\gamma_1^2d_2P\Big)\Big\} \end{aligned} \tag{3.53}$$

and the diffusion coefficient $a(P)$ is

$$\begin{aligned} \frac{1}{2}a(P)^2 &= (\frac{\mathrm{d}P}{\mathrm{d}\Pi_1})^2\Big(d_1\pi_1(\mathrm{d}_1\pi_1)^2 + d_2\pi_2(\mathrm{d}_2\pi_1)^2\Big) \\ &= \frac{\mathrm{d}P}{\mathrm{d}\Pi_1}\frac{P(1-P)\kappa_2}{\Big(\gamma_1\kappa_1p + \gamma_2\kappa_2(1-P)\Big)^2}\Big(\kappa_2(1-P)+\kappa_1P\Big)\Big(\gamma_2^2d_1(1-P)+\gamma_1^2d_2P\Big), \end{aligned} \tag{3.54}$$

with $d_k = D + \mu_k$, $\gamma_k = \frac{(m_k-d_k)^2}{m_ka_k}$, $\kappa_k = \frac{d_k}{y_k}$.

In this section, by applying Itô's transformation and the weak convergence, we derived an explicit diffusion approximation for the stochastic process at a long time scale, whose behaviour is on the centre manifold: $\kappa_1\Pi_1 + \kappa_2\Pi_2 = (S^{\mathrm{in}} - S^*)D$, until one of the absorption states, $\Pi_1 = 0$

or $\Pi_2 = 0$ is reached. This is the classical splitting problem in stochastic analysis, which is defined as a fixation problem in ecology with the changing from the situation of multiple species coexistence to dominance by a single species.[4]

In the following section, the fixation probabilities and the mean of the first absorption time will be calculated.

3.7 Fixation Problems

The stochastic fluctuations on the centre manifold drive one population to extinction and as the cause of the weak competition. The absorbing states are where $\Pi_1 = 0$ or $\Pi_2 = 0$ (one species dominates the whole population). The splitting probability on the centre manifold to one of the absorbing states is also the fixation probability in population genetics which is a useful quality to describe the competitive ability of a quasi-neutral species by comparing with the strictly neutral model.

In Ewens's book [16], he demonstrates how to derive the fixation probabilities and the absorption time properties. Using the backward Kolmogorov equation, the partial differential equation (PDE) of the relative abundance of the first species $P(t)$ in the future time t will reach

$$\frac{\mathrm{d}f(x,t;\ p)}{\mathrm{d}t} = b(p)\frac{\mathrm{d}f(x,t;\ p)}{\mathrm{d}p} + \frac{1}{2}a(p)^2\frac{\mathrm{d}^2f(x,t;\ p)}{\mathrm{d}p^2}$$

with the initial state $P(0) = p$. The random variable x is the value of the

[4] In genetics, fixation is the change in a gene pool from a situation where there exists at least two variants of a particular gene (allele) to a situation where only one of the alleles remains.

relative abundance of the first species at time t, and $b(p), a(p)^2$ are expressed in Eqs. (3.53) and (3.54).

The points $x = 0$ and 1 are the absorbing states of the diffusion process. Define the fixation probability of the first species with initial relative abundance p as $F(p)$, which is the probability that the first species dominates the population at or before infinite time t ($t \to \infty$). Define the mean time until either of the absorbing boundaries is reached as $ET(p)$. Ewens discusses these quantities in [15] for general cases, and gives their ODEs with boundary conditions. In this section, we will apply Ewens's methods to derive the fixation quantities in our model.

Before proceeding to the next analytical step, it is necessary to clearly state the coefficients of the diffusion. $b(p)$ and $a(p)$ are already presented in Eqs. (3.53) and (3.54), and yield

$$\begin{aligned}
\frac{b(p)}{\frac{1}{2}a(p)^2} &= \frac{\gamma_1\gamma_2(\kappa_2 d_2 - \kappa_1 d_1)}{\Big(\kappa_1 p + \kappa_2(1-p)\Big)\Big(\gamma_1^2 d_2 p + \gamma_2^2 d_1(1-p)\Big)} \\
&+ \frac{\gamma_1\gamma_2(\gamma_1\kappa_2 d_2 - \gamma_2\kappa_1 d_1)}{\Big(\kappa_1\gamma_1 p + \kappa_2\gamma_2(1-p)\Big)\Big(\gamma_1^2 d_2 p + \gamma_2^2 d_1(1-p)\Big)} + 2\frac{\kappa_1 - \kappa_2}{\kappa_1 p + \kappa_2(1-p)} \\
&= \Big(2 + \frac{\gamma_1\gamma_2(\kappa_2 d_2 - \kappa_1 d_1)}{\gamma_2^2\kappa_1 d_1 - \gamma_1^2\kappa_2 d_2}\Big)\frac{\kappa_1 - \kappa_2}{(\kappa_1 - \kappa_2)p + \kappa_2} \\
&- \frac{(\gamma_1\kappa_2 d_2 - \gamma_2\kappa_1 d_1)(\gamma_1 + \gamma_2)}{\gamma_2^2\kappa_1 d_1 - \gamma_1^2\kappa_2 d_2}\frac{d_2\gamma_1^2 - d_1\gamma_2^2}{(d_2\gamma_1^2 - d_1\gamma_2^2)p + d_1\gamma_2^2} \\
&- \frac{\gamma_1\kappa_1 - \gamma_2\kappa_2}{(\gamma_1\kappa_1 - \gamma_2\kappa_2)p + \gamma_2\kappa_2}
\end{aligned} \tag{3.55}$$

where $\kappa_k = \frac{d_k}{y_k}$.

Now, we will use these results to calculate the fixation probability of the first species and the mean of the first absorption time.

3.7.1 Fixation Probability of the First Species $F(p)$

As Ewens shows in [16], the fixation probability $F(p)$ of the first species satisfies the ODE system:

$$b(p)\frac{dF}{dp} + \frac{a(p)^2}{2}\frac{d^2F}{dp^2} = 0 \tag{3.56}$$

$$F(p=0) = 0, \tag{3.57}$$

$$F(p=1) = 1. \tag{3.58}$$

Firstly, by taking the integral with respect to p over Eq. (3.56), we have

$$\begin{aligned}\frac{dF}{dp} &= c_1 e^{-2\int \frac{b(p)}{a(p)^2}dp} \\ &= c_1\psi(p).\end{aligned}$$

With the conditions Eqs. (3.57) and (3.58), the solution is then

$$F(p) = \frac{\int_0^p \psi(u)du}{\int_0^1 \psi(u)du}, \tag{3.59}$$

where $\psi(p)$ is expressed as:

$$\psi(p) = \frac{(\gamma_1\kappa_1 - \gamma_2\kappa_2)p + \gamma_2\kappa_2}{[(\kappa_1-\kappa_2)p+\kappa_2]^3}\left\{\frac{(d_2\gamma_1^2 - d_1\gamma_2^2)p + d_1\gamma_2^2}{(\kappa_1-\kappa_2)p+\kappa_2}\right\}^{\frac{(\gamma_1\kappa_2 d_2 - \gamma_2\kappa_1 d_1)(\gamma_1+\gamma_2)}{\gamma_2^2\kappa_1 d_1 - \gamma_1^2\kappa_2 d_2}}.$$

This is a general results for the fixation probability. Furthermore, as the intrinsic death rates are assumed to be insignificant, the death rates of the

species are identical, i.e., $d_1 = d_2 = D$, and $\psi(p)$ is simplified as

$$\psi(p) = \frac{(\gamma_1\kappa_1 - \gamma_2\kappa_2)p + \gamma_2\kappa_2}{\Big((\kappa_1 - \kappa_2)p + \kappa_2\Big)^3} \left\{ \frac{(\gamma_1^2 - \gamma_2^2)p + \gamma_2^2}{(\kappa_1 - \kappa_2)p + \kappa_2} \right\}^{\frac{(\gamma_1\kappa_2 - \gamma_2\kappa_1)(\gamma_1 + \gamma_2)}{\gamma_2^2\kappa_1 - \gamma_1^2\kappa_2}}. \quad (3.60)$$

Unfortunately, this fixation probability is too complicated to derive a general results of the effects of the parameters. The sign of $b(p)$ which explains the direction of the flow on the centre manifold, is decided by six parameters $\gamma_k, \kappa_k (= \frac{d_k}{y_k}), m_k$ and the initial relative abundance p. We will discuss the effects of these parameters in specific conditions in the following subsections.

3.7.1.1 Neutral Case

To verify the result in the case of the strictly neutral model, let $\gamma_1 = \gamma_2$, $d_1 = d_2$, $y_1 = y_2$, we have

$$\psi(p) = 1.$$

Then the fixation probability of the first species in the neutral model is the reciprocal of the initial frequency by symmetry, that is

$$F(p) = p.$$

Each neutral species has an equal chance of being the one that will eventually become fixed via drift. Therefore, by comparing with this symmetric result, we may demonstrate a selective advantage for either of the species in the quasi-neutral model.

3.7.1.2 Effects of γ_k With Identical y_k

To determine how the γ_k may affect the fixation probability without the impact of the yields, we assume the yields to be identical. Since $\kappa_k = \frac{d_k}{y_k}$, with the assumption of $d_k = D$, we have $\kappa_1 = \kappa_2$. Then

$$\begin{aligned}
\psi(p) =& \frac{(\frac{\gamma_1}{\gamma_2} - 1)p + 1}{(\frac{\gamma_1}{\gamma_2}^2 - 1)p + 1} \\
F(p) =& \frac{\int_0^p \psi(u)\mathrm{d}u}{\int_0^1 \psi(u)\mathrm{d}u} = \frac{p + \frac{\frac{\gamma_1}{\gamma_2}}{\frac{\gamma_1}{\gamma_2}^2 - 1} \log((\frac{\gamma_1}{\gamma_2}^2 - 1)p + 1)}{1 + \frac{\frac{\gamma_1}{\gamma_2}}{\frac{\gamma_1}{\gamma_2}^2 - 1} \log\{\frac{\gamma_1}{\gamma_2}^2\}} \\
=& p + \frac{\frac{\gamma_1}{\gamma_2}[\log\Big((\{\frac{\gamma_1}{\gamma_2}^2\} - 1)p + 1\Big) - 2p\log\frac{\gamma_1}{\gamma_2}]}{\frac{\gamma_1}{\gamma_2}^2 - 1 + 2\frac{\gamma_1}{\gamma_2}\log\frac{\gamma_1}{\gamma_2}}.
\end{aligned}$$

That is to say,

$$F(p) > p, \text{ if } \gamma_1 > \gamma_2,$$
$$F(p) < p, \text{ if } \gamma_1 < \gamma_2.$$

This result helps us to understand the impact of γ_k when the yields are fixed, the species with higher γ_k has an increased probability of surviving the weak competition, and dominating the whole population.

3.7.1.3 Effects of y_k With Identical γ_k

If we assume the species share the same trade-off parameter $\gamma_k = (\frac{\mathrm{d}b_k(S)}{\mathrm{d}S})|_{S^*}$, then the a_k and m_k are identical for the quasi-neutral species. The fixation

probability is obtained as

$$\psi = 1,$$
$$F(p) = p,$$

which is exactly the same as the fixation probability in the strictly neutral model.

Therefore, if $\gamma_1 = \gamma_2$, the yields do not impact the fixation probability in the weak competition. However, they will affect the population size and the time to the first absorption state.

3.7.2 Corrections to the Fixation Probability

Above we presented the analytical calculation for the fixation probabilities. However, in the next chapter, when comparing these analytical predictions with simulation results, we will find large discrepancies. Moreover, these difference will not disappear as the simulation population increases.

This could be explained by the fact that, in the derivation of the diffusion approximation, we use the first order Taylor expansion of the deterministic system to derive the derivatives of the projection map. The absence of the higher order terms of the Taylor expansion in the calculations leads the final analytical fixation probability not being accurate. Therefore, in this subsection, we take the higher order term of the Taylor expansion of the deterministic trajectory into account, and use the corrected derivatives of the projection map (derived in Section 3.5.3.1) to express the diffusion approximation. Correspondingly, the drift coefficient $b(P)$ in the diffusion

approximation of the relative abundance will be corrected to a new value with notation $b_1(P)$:

$$b_1(P) = b(P) + \frac{\mathrm{d}P}{\mathrm{d}\Pi_1}\frac{D^2}{S^*}\frac{\kappa_2\gamma_1\gamma_2 P(1-P)}{\Big((\kappa_1\gamma_1-\kappa_2\gamma_2)P+\kappa_2\gamma_2\Big)^3}(\frac{1}{m_2}-\frac{1}{m_1})\Big((d_1\kappa_1^2-d_2\kappa_2^2)P+d_2\kappa_2^2\Big).$$

The diffusion coefficient $a(P)$ (expressed in Eq. (3.54)) is unchanged because it is only dependent on the first derivatives of the projection map which are kept constant.

Then, the coefficient $\frac{b_1(p)}{\frac{1}{2}a(p)^2}$ can be obtained,

$$\frac{b_1(p)}{\frac{1}{2}a(p)^2} = \frac{b(p)}{\frac{1}{2}a(p)^2} + \frac{\gamma_1\gamma_2 D^2}{S^*}\Big(\frac{1}{m_2}-\frac{1}{m_1}\Big)\frac{(\kappa_1^2d_1-\kappa_2^2d_2)p+\kappa_2^2d_2}{(\gamma_1^2d_1-\gamma_2^2d_1)p+\gamma_2^2d_1}\frac{1}{(\kappa_1-\kappa_2)p+\kappa_2}\frac{1}{(\gamma_1\kappa_1-\gamma_2\kappa_2)p+\gamma_2\kappa_2} \tag{3.61}$$

Applying this new results of $\frac{b_1(p)}{\frac{1}{2}a(p)^2}$ in backward Kolmogrove equation, we have a corrected fixation probability expressed as:

$$F(p) = \frac{\int_0^p \psi_1(u)du}{\int_0^1 \psi_1(u)du}$$

with

$$\begin{aligned}\psi_1(p) =&\{(\gamma_1\kappa_1-\gamma_2\kappa_2)p+\gamma_2\kappa_2\}^{1-\frac{D^2}{(\gamma_1-\gamma_2)S^*}(\frac{1}{m_2}-\frac{1}{m_1})}\\ &\{(\gamma_1^2-\gamma_2^2)p+\gamma_2^2\}^{\frac{(\gamma_1\kappa_2-\gamma_2\kappa_1)(\gamma_1+\gamma_2)}{\gamma_2^2\kappa_1-\gamma_1^2\kappa_2}-\frac{\kappa_2^2\gamma_1^2-\kappa_1^2\gamma_2^2}{(\kappa_2\gamma_1^2-\kappa_1\gamma_2^2)(\gamma_1\kappa_2-\gamma_2\kappa_1)}\frac{D^2}{(\gamma_1-\gamma_2)S^*}(\frac{1}{m_2}-\frac{1}{m_1})}\\ &\{(\kappa_1-\kappa_2)p+\kappa_2\}^{-3-\frac{(\gamma_1\kappa_2-\gamma_2\kappa_1)(\gamma_1+\gamma_2)}{\gamma_2^2\kappa_1-\gamma_1^2\kappa_2}-\frac{\kappa_1-\kappa_2}{(\gamma_1-\gamma_2)(\gamma_1^2\kappa_2-\gamma_2^2\kappa_1)}\frac{D^2}{(\gamma_1-\gamma_2)S^*}(\frac{1}{m_2}-\frac{1}{m_1})}\end{aligned} \tag{3.62}$$

3.7.3 Mean of the First Absorption Time $ET(p)$

As explained in [16], the mean of the first absorption time is the solution of

$$\begin{aligned} b(p)\frac{dET}{dp}+\frac{a(p)^2}{2}\frac{d^2ET}{dp^2} &= -1,\\ ET(p=0) &= 0,\\ ET(p=1) &= 0,\end{aligned}$$

with the result in the form of

$$ET(p)=\int_0^1 t(x;p)\mathrm{d}x$$

where

$$\begin{aligned} t(x;p) &= 2\frac{1-F(p)}{a^2(x)\psi(x)}\int_0^x \psi(y)\mathrm{d}y, \quad 0\le x\le p\\ t(x;p) &= 2\frac{F(p)}{a^2(x)\psi(x)}\int_x^1 \psi(y)\mathrm{d}y, \quad p\le x\le 1.\end{aligned}$$

Then the mean of the first absorption time can be expressed as,

$$ET(p) = 2\int_0^1 \psi(x)\mathrm{d}x \cdot \Big\{(1-F(p))\int_0^p \frac{F(x)}{a^2(x)\psi(x)}\mathrm{d}x + F(p)\int_p^1 \frac{1-F(x)}{a^2(x)\psi(x)}\mathrm{d}x\Big\}. \tag{3.63}$$

This expression is even more complicated than the fixation probability, and it is impossible to directly determine its sensitivity to each parameter. In the next chapter, we will plot this function with different parameter values to determine how the times change as the parameters vary, and to compare them with the numerical results.

Neutral Case

In the case of the truly neutral model, by assuming $\gamma_1 = \gamma_2$, $d_1 = d_2$, $y_1 = y_2$, the mean of the first absorption time is simplified as

$$ET(p) = -y(S^{\text{in}} - S^*)\Big(p\log p + (1-p)\log(1-p)\Big).$$

As the population size Θ equals $y(S^{\text{in}} - S^*)$ at the equilibrium state, we have

$$ET(p) = \Theta^*\Big(-p\log p - (1-p)\log(1-p)\Big).$$

Therefore, in the strictly neutral case, the mean of the first absorption time is proportional to the yield through the impact on the equilibrium population size.

3.8 Quasi-Stationary Distribution

We have explained in the previous sections that the finite Markov process will eventually be trapped into the extinction state. In the microbial world, a huge population will endure for a very long time before dying out far beyond the human timescale. Therefore it is interesting to find out the information contained in this long time process before the predestined absorption.

In an island community with a population that has immigrated from a meta-community, if the relative abundance of each species in the immigration flow is positive and constant, then the population on the island community will have coexistence stationary states, that means the stationary distribution of the relative abundance exists. This will be discussed in Chapter 5.

In our current isolated chemostat model, once the first absorption state is reached, there is only one species left in the whole population, and it will endure for a long time before fading away. Due to the final extinction, there is no strictly stationary distribution for the abundance of the single species that dominates the whole population after the weak selection, but there is an asymptotically stationary distribution prior to this final extinction. This is defined as the quasi-stationary distribution, which is an important probabilistic measure of the system behaviour. Correspondingly, the phase between the first absorption being reached and the final extinction is defined as the quasi-stationary phase.

In this section, we will first determine this asymptotically stationary distribution of the surviving species if the condition of non-extinction is

assumed. Then we will discuss whether a quasi-stationary distribution exists for the relative abundance given the condition that the absorbing state has not been reached.

3.8.1 Quasi-Stationary Distribution of the First Species $f_{q,1}$

We can assume without loss of generality that it is the first species that fixes in the whole population in the quasi-stationary phase where there are no individuals of the second species. The quasi-stationary distribution $f_{1,q}(\hat{x}_1)$ is a conditional probability density with definition,

$$f_{q,1}(x) = \lim_{t\to\infty} \Pr\Big(\hat{x}_1(t) \in \lim_{\Delta x\to 0}(x, x+\Delta x) \mid \hat{x}_1(t) \neq 0\Big),$$

which is independent of the time t after the first absorption occurs.

In this subsection, the approach by Van der Werf [45] will be used to derive this conditional distribution in the quasi-stationary phase. As discussed in Subsection 3.4.1, with the help of the central limit theorem, the centred and scaled stochastic process $\sqrt{V}(\hat{x}(t) - \bar{x}(t))$ was proved to be approximated by an Ornstein-Uhlenbeck process $X(t)$,

$$\mathrm{d}X(t) = \partial F\Big(\bar{x}(t)\Big)X(t)\mathrm{d}t + \sum_l l\beta_l^{\frac{1}{2}}\Big(\bar{x}(s)\Big)\mathrm{d}W_l.$$

By finding out the quasi-stationary distribution of $X(t)$, we can determine the $f_{q,1}(x)$.

Since the second species has already died out, the process is now a bi-

variate Ornstein-Uhlenbeck system,

$$\mathrm{d}X(t) = -\Theta(X - E)\mathrm{d}t + C\mathrm{d}B_t.$$

with

$$\begin{aligned} E &= 0, \\ \Theta &= \begin{pmatrix} D + \frac{\gamma_1}{y_1}x^* & \frac{d_1}{y_1} \\ -\gamma_1 x^* & 0 \end{pmatrix}, \\ \mathbf{G} &= CC' = \begin{pmatrix} 0 & 0 \\ 0 & 2d_1 x_1^* \end{pmatrix}. \end{aligned}$$

As Van der Werf (2007) discusses in [45], the conditional distribution of the Ornstein-Uhlenbeck process is normally distributed at all times,

$$X_{t+\tau} \approx N(x_{t+\tau}, \sum_{\tau}),$$

with the deterministic drift

$$x_{t+\tau} = (I - e^{-\Theta\tau})E + e^{-\Theta\tau}x_t,$$

and the covariance $\sum_\tau$ which is expressed in terms of the stack operator

vec[5] and the Kronecker sum $\oplus$[6],

$$vec(\sum_{\tau}) = (\Theta \oplus \Theta)^{-1}(I - e^{-(\Theta \oplus \Theta)\tau})vec(G).$$

As τ goes to infinity, since the eigenvalue of Θ has positive real part, the

[5]Operator vec: The vectorization of a matrix is a linear transformation which converts the matrix into a column vector. Specifically, the vectorization of an $m \times n$ matrix A, denoted by $vec(A)$, is the $mn\times$ 1 column vector obtained by stacking the columns of the matrix A on top of one another:

$$\text{vec}(A) = [a_{1,1}, \ldots, a_{m,1}, a_{1,2}, \ldots, a_{m,2}, \ldots, a_{q,n}, \ldots, a_{m,n}]^T$$

[6]Kronecker sum $\oplus$: If A is an $m \times n$ matrix and B is a $p \times q$ matrix, then the Kronecker project $A \oplus B$ is the $mp \times nq$ block matrix:

$$A \oplus B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{pmatrix}$$

distribution of the process stabilises to a Gaussian distribution with

$$
\begin{aligned}
x_\infty &= E, \\
vec(\sum_\infty) &= (\Theta \oplus \Theta)^{-1} vec(G) \\
&= \begin{pmatrix} 2(D+\frac{\gamma_1}{y_1}x_1^*) & \frac{d_1}{y_1} & \frac{d_1}{y_1} & 0 \\ -\gamma_1 x_1^* & D+\frac{\gamma_1}{y_1}x_1^* & 0 & \frac{d_1}{y_1} \\ -\gamma_1 x_1^* & 0 & D+\frac{\gamma_1}{y_1}x_1^*\frac{d_1}{y_1} & \frac{d_1}{y_1} \\ 0 & -\gamma_1 x_1^* & -\gamma_1 x_1^* & 0 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2dx_1^* \end{pmatrix}, \\
&= \begin{pmatrix} \frac{1}{2(D+\frac{\gamma_1}{y_1}x_1^*)} & 0 & 0 & \frac{\frac{d_1}{y_1}}{2(D+\frac{\gamma_1}{y_1}x_1^*)\gamma_1 x_1^*} \\ 0 & \frac{1}{2(D+\frac{\gamma_1}{y_1}x_1^*)} & -\frac{1}{2(D+\frac{\gamma_1}{y_1}x_1^*)} & -\frac{1}{2\gamma_1 x_1^*} \\ 0 & -\frac{1}{2(D+\frac{\gamma_1}{y_1}x_1^*)} & \frac{1}{2(D+\frac{\gamma_1}{y_1}x_1^*)} & -\frac{1}{2\gamma_1 x_1^*} \\ \frac{\gamma_1 x_1^*}{2(D+\frac{\gamma_1}{y_1}x_1^*)\frac{d_1}{y_1}} & \frac{y_1}{2d_1} & \frac{y_1}{2d_1} & \frac{D+\frac{\gamma_1}{y_1}x_1^*}{2\frac{d_1}{y_1}\gamma_1 x_1^*} + \frac{1}{2(D+\frac{\gamma_1}{y_1}x_1^*)} \end{pmatrix} \\
&\quad \times \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2d_1 x_1^* \end{pmatrix}, \\
&= \begin{pmatrix} \frac{d_1^2}{(Dy_1+\gamma_1 x_1^*)\gamma_1} \\ -\frac{d_1}{\gamma_1^*} \\ -\frac{d_1}{\gamma_1^*} \\ \frac{y_1(D+\frac{\gamma_1}{y_1}x_1^*)}{\gamma_1} + \frac{x_1^* d_1}{D+\frac{\gamma_1}{y_1}x_1^*} \end{pmatrix}.
\end{aligned}
$$

After being transformed by Eq. (3.14), the conditional probability den-

sity of the first species is also Gaussian distributed with mean and variance,

$$E_1 = x_1^* = \frac{y_1(S^{\text{in}} - S^*)D}{d_1} \tag{3.64}$$

$$\sigma_1^2 = \frac{1}{V}\left(\frac{y_1(D + \frac{\gamma_1}{y_1}x_1^*)}{\gamma_1} + \frac{x_1^* d_1}{D + \frac{\gamma_1}{y_1}x_1^*}\right). \tag{3.65}$$

This is the quasi-stationary distribution of the first species.

Quasi-Stationary Distribution of the Second Species

Similarly, if the second species fixes in the whole population in the quasi-stationary phase, the quasi-stationary distribution for the density of the second species is

$$f_{q,2}(x) = \lim_{t\to\infty} \Pr\Big(\hat{x}_2(t) \in \lim_{\Delta x\to 0}(x, x + \Delta x) \mid \hat{x}_2(t) \neq 0\Big),$$

which is Gaussian distributed with mean and variance,

$$E_2 = \frac{y_2(S^{\text{in}} - S^*)D}{d_2}$$

$$\delta_2^2 = \frac{1}{V}\left(\frac{y_2(D + \frac{\gamma_2}{y_2}x_2^*)}{\gamma_2} + \frac{x_2^* d_2}{D + \frac{\gamma_2}{y_2}x_2^*}\right).$$

However, in the coexistence phase over compact time intervals, the stochastic processes with two quasi-neutral species does not behave as a Gaussian distribution due to the existence of the zero eigenvalue. The existence problem of the quasi-stationary distribution of the relative abundance

$f_q(P(t) = x)$ will be discussed in Section 3.8.3.

Moreover, as illustrated in the Figure 2.4 (in Section 2.3), when the deterministic dynamics returns to any point on the flows with direction towards the state (x_1^*, x_2^*), it will always move back to (x_1^*, x_2^*). Thus for a specific state (x_1^*, x_2^*) on the centre manifold, there is a stationary distribution along the flow with direction towards this state. The method of calculation is similar to that in Subsection 3.8.1, but more tedious for a stationary distribution of a trivariate Ornstein-Uhlenbeck system.

3.8.2 Relationship Between the Quasi-Stationary Distributions and the Fixation Probabilities

It is well known that, given a smaller variance, the stochastic process is more stable and will remain closer to the equilibrium state. Conversely, a species with larger variance has larger fluctuations, and is more likely to reach the extinction state. In our chemostat model, if $V \to \infty$, the variance is close to zero, the stochastic processes is stable remaining near the equilibrium state, then the quasi-stationary phase will be very long. Given the explicit expressions of the quasi-stationary distributions and the fixation probabilities, we wonder whether there is a certain relationship between them? Does a more stable Gaussian distributed species have an advantage in terms of fixation probability before this monotonic state is reached? These questions will be answered in this subsection.

To compare these two quasi-neutral species, we list their distribution

coefficients:

$$var_1 = \frac{1}{V}\{D(\frac{x_1^*}{D} + \frac{y_1}{\gamma_1}) + \frac{D}{\frac{D}{x_1^*} + \frac{\gamma_1}{y_1}}\} \tag{3.66}$$

$$x_1^* = (S^{\text{in}} - S^*)y_1 \tag{3.67}$$

$$var_2 = \frac{1}{V}\{x_2^* + \frac{y_2 D}{\gamma_2} + \frac{D}{\frac{D}{x_2^*} + \frac{\gamma_2}{y_2}}\} \tag{3.68}$$

$$x_2^* = (S^{\text{in}} - S^*)y_2 \tag{3.69}$$

where $\gamma_k = \frac{\partial b_k(S)}{\partial S}|_{S^*}$.

Here are the connections which we find between the quasi-stationary distributions and the fixation probabilities:

- In the strictly neutral case, the two species are ecologically identical, thus the fixation probability is decided only by the initial states.

- If the yields y_k are identical, the two species share the same mean value in a single species environment, i.e., $x_1^* = x_2^*$. The larger γ_k corresponds to a smaller variance, that gives the species a more stable population. As concluded in Subsection 3.7.1.2, the species with larger γ_k has an increased fixation probability. The species with an advantage during weak selection is associated with a more stable Gaussian process in the quasi stationary phase, and *vice versa*.

- If we let $\gamma_1 = \gamma_2$, the mean and the variance of the quasi-stationary distributions in Eqs. (3.66)-(3.69) are dependent on the yields y_k, but

not on their ratios which are identical as,

$$\frac{var_1}{x_1^*} = \frac{var_2}{x_2^*} = \frac{1}{V}\{1 + \frac{D}{\gamma(S^{\text{in}} - S^*)} + \frac{D}{D + \gamma(S^{\text{in}} - S^*)}\}.$$

Consequently these two Gaussian processes have equivalent strengths of relative fluctuations. Applying $\gamma_1 = \gamma_2$ to the fixation probability, we get the neutral result $F(p) = p$ in Subsection 3.7.1.3. Therefore, quasi-neutral species that share the same γ_k will have the same dispersal of their Gaussian distribution, and act ecologically identically in the weak competition, no matter how different their yields are.

3.8.3 Distribution of the Relative Abundance in the Coexistence Phase

3.8.3.1 Does a Quasi-Stationary Distribution of the Relative Abundance Exis?

The quasi-stationary distribution of a single species after the long coexistence phase is well understood, now we will address the question of whether there exists an asymptotically stationary distribution for the relative abundance in the long-time coexistence phase? Is there a quasi stationary regime that the coexisting populations could be relaxed to?

The answer is no. The conditional distribution of the Ornstein-Uhlenbeck process

$$\mathrm{d}X(t) = -\Theta(X - E)\mathrm{d}t + C\mathrm{d}B_t.$$

is stationary only if the real parts of all the eigenvalues of Θ are strictly

positive. In the coexistence phase of our model, zero eigenvalues exists. Therefore the overall covariance of the Ornstein-Uhlenbeck process does not converge and no quasi-stationary distribution of relative abundance exist in the coexistence stage of the model. The probability density of the relative abundance is then time-dependent.

3.8.3.2 Probability Density of the Relative Abundances

If the population is large, then the duration time until the first absorption will be long enough to consider the behaviour of the relative abundances, even though its distribution is not quasi-stationary. Before proceeding, it is necessary to issue some caveats and define some terms.

First of all, the probability density of the relative abundance of the first species conditional on no absorption is defined as

$$f_c(x,t|p,0) = \lim_{\Delta x \to 0} \Pr\Big(P_t \in (x, x+\Delta x) \mid P_t \in (0,1),\ P_0 = p\Big). \tag{3.70}$$

Secondly, define $f(x,t \mid p,0)$ as the probability density of the relative abundance of the first species with definition

$$f(x,t|p,0) = \lim_{\Delta x \to 0} \Pr\Big(P_t \in (x, x+\Delta x) \mid P_0 = p\Big). \tag{3.71}$$

To distinguish, $f_c(x,t|p,0)$ is denoted as conditional probability density and $f(x,t|p,0)$ as a probability density. In this section, we will determine the relationship between the probability density $f(x,t \mid p,0)$ and the conditional probability density $f_c(x,t \mid p,0)$.

Thirdly, we define $v(p,t)$ as the probability that the whole population is not monotypic at time t, with expression,

$$\begin{aligned} v(p,t) &= \Pr(P_t \in (0,1)|P_0 = p) \\ &= \Pr(\text{Time to the first absorption state} > t|P_0 = p) \\ &= 1 - \Pr(\text{Time to the first absorption state} \leq t|P_0 = p) \end{aligned}$$

In almost all density-dependent population dynamics models, the time to extinction is exponentially distributed [25] [48]. Thus using the known expectation of the first absorption time $ET(p)$ in Eq. (3.63), $v(p,t)$ can be determined as,

$$v(p,t) = 1 - \int_0^t \frac{1}{ET(p)} \exp\{-\frac{s}{ET(p)}\}\mathrm{d}s \tag{3.72}$$

We already know that both the conditional and unconditional probability density of the relative abundance are dependent on t. To find their relationships, we may apply the Bayes' theorem on the conditional density

Eq. (3.70),

$$\begin{aligned} f_c(x,t) &= \lim_{\Delta x \to 0} \Pr\Big(P_t \in (x, x+\Delta x) \mid P_t \in (0,1),\ P_0 = p\Big) \\ &= \lim_{\Delta x \to 0} \frac{\Pr\Big(P_t \in (x, x+\Delta x),\ P_t \in (0,1) \mid P_0 = p\Big)}{\Pr\Big(P_t \in (0,1) \mid P_0 = p\Big)} \\ &= \lim_{\Delta x \to 0} \frac{\Pr\Big(P_t \in (x, x+\Delta x) \mid P_0 = p\Big) \cdot \Pr\Big(P_t \in (0,1) \mid P_t \in (x, x+\Delta x), P_0 = p\Big)}{\Pr\Big(P_t \in (0,1) \mid P_0 = p\Big)} \\ &= f(x,t \mid p,0)\ \frac{\Pr\Big(P_t \in (0,1) \mid P_t \in (x, x+\Delta x), P_0 = p\Big)}{\Pr\Big(P_t \in (0,1) \mid P_0 = p\Big)} \end{aligned}$$

where

$$\Pr\Big(P_t \in (0,1) \mid P_t \in (x, x+\Delta x), P_0 = p\Big) = \begin{cases} 1 & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases} \tag{3.73}$$

is the indicator function $\mathbb{1}_{\{x\in(0,1)\}}$.

Thus $f_c(x,t)$ could be derived with $f(x,t|p,0)$ and $v(p,t)$, that,

$$f_c(x,t) = \frac{f(x,t \mid p,0)}{v(p,t)}\ \mathbb{1}_{\{x\in(0,1)\}} \tag{3.74}$$

where $v(x,t)$ is expressed in Eq. (3.72). Using the expression of the diffusion process

$$\mathrm{d}P_t = b(P_t)\mathrm{d}t + a(P_t)\mathrm{d}t,$$

and its initial state $P(t=0) = p$, the probability density $f(x,t|p,0)$ in the

coexistence stage satisfies the Fokker-Planck equation,

$$\frac{\mathrm{d}f(x,t|p,0)}{\mathrm{d}t} = -\frac{\mathrm{d}b(x)f(x,t)}{\mathrm{d}x} + \frac{1}{2}\frac{\mathrm{d}^2a(x)^2f(x,t)}{\mathrm{d}x^2} \tag{3.75}$$

Therefore, the conditional probability density of the relative abundance of the first species $f_c(x,t)$ is the solution of the system:

$$\frac{\mathrm{d}f(x,t|p,0)}{\mathrm{d}t} = -\frac{\mathrm{d}b(x)f(x,t|p,0)}{\mathrm{d}x} + \frac{1}{2}\frac{\mathrm{d}^2a(x)^2f(x,t|p,0)}{\mathrm{d}x^2} \tag{3.76}$$

$$f_c(x,t) = f(x,t \mid p,0)\exp\Big\{-\frac{1}{ET(p)}t\Big\}\ \mathbb{1}_{\{x\in(0,1)\}} \tag{3.77}$$

with $ET(x)$ and the parameters $b(x)$, $a(x)$ derived in Eqs. (3.63), (3.53) and (3.54) respectively.

Although a generally analytical solution of the partial differential equation Eq. (3.76) is impossible to derive, it can be calculated numerically with simple parameter assumptions to predict the value of $f_c(x,t)$ through Eq. (3.77).

3.9 Summary

As discussed in the deterministic model, after competition, we have a strong selection regime in the equilibrium state. The species that share the largest deterministic fitness win the competition and coexist on the centre manifold indefinitely through the life history trade-offs. This is the niche assembly rule.

In this chapter, we incorporated stochastic drift into the dynamics, and focussed on the stochastic behaviours of these quasi-neutral species on the

centre manifold, over compact time intervals and infinite time intervals. Over compact time intervals, the law of large numbers and the central limit theorem both work well. The stochastic dynamics closely fluctuates around the deterministic dynamics, and may approximate to an Ornstein-Uhlenbeck process. In all real microbial communities, the population is large but finite, then over an infinite interval, the law of large numbers and the central limit theorem are no longer applicable. Weak competition exists among the quasi-neutral species, and leads the coexistence states on the centre manifold to one of the absorbing states, where a single species dominates the whole population. Therefore, a dominant role played by stochastic drift in the long time dynamics can be determined.

In order to derive the diffusion approximation over an infinite time interval, we changed the time scale of the stochastic process, and then found that the projection map onto the deterministic process as $t \to \infty$ can well approximate the stochastic process at a long time scale. Using Itô's transformation and weak convergence, the diffusion approximation of the stochastic process at a long time scale was derived. Using this specific diffusion approximation, all quantities regarding the fixation problem may be analytically calculated. The stationary distribution in the quasi-stationary stage is another important idea in the population dynamics, which is not only Gaussian but also found to be related to the fixation probability. In the absence of immigration, the quasi-stationary distribution of the relative abundance does not exist in the model as it possesses a zero eigenvalue.

This chapter explains the significant role stochastic drift plays in a compact microbial community in the absence of immigration, especially over

infinite time intervals, demonstrating that random effects do lead to extinction in scenarios where the deterministic model predicts persistence. Even tiny stochastic drift will project the stochastic process onto a different deterministic limiting result after a long enough time, and change the original deterministic coexistence state.

Both in the deterministic and stochastic models, the trade-off parameter γ_k which measures the responsiveness of the microbial growth rate to changes in the substrate concentration at the equilibrium value, is a useful parameter that helps us predict the dynamics of the process. In the case of deterministic strong selection, the species with large γ_k has a larger birth rate until the equilibrium is reached, and thus this species will have a larger population abundance in the coexistence equilibrium state. In the stochastic weak selection, larger γ_k is also an advantage which helps the species dominate the whole population. If the γ_k are identical, the quasi-neutral species appear truly neutral, no matter how different their yields are. Therefore, the quasi-neutral species with different trade-off parameters exhibits substantial differences even with the same largest fitness. The mechanism of the life history trade-offs could not ensure the coexistence and equivalence of the quasi-neutral species in this model with the effects of the stochastic drift.

To balance extinction and maintain the population diversity, it is necessary to introduce immigration into the model. This new model will be analysed in Chapter 5. Before that, in the next chapter, the analytical results of the strong selection, weak selection and the quasi-stationary distributions will be checked by comparing them with the numerical calculations for small

populations.

Chapter 4

Effects of the Parameters: Comparison of the Analytical Model and the Numerical Model

From the analytical discussions in the last two chapters, we know that if the volume V (in the units of the volume of a typical cell) in a chemostat is finite but close to infinitely large, the continuous stochastic process converges to the deterministic dynamics over finite time intervals and behaves as an Ornstein-Uhlenbeck process. Over an infinite time interval, a new diffusion approximation is analytically calculated, from which we can predict the fixation probability and the mean of the first absorption time. Furthermore, the best approach to depict the dynamics of the stochastic

model and to check the accuracy of the analytical diffusion approximation, is through numerical simulation. In this chapter, we will simulate the discrete stochastic model in the absent of immigration for small population, and compare the results of the strong and weak selection with the analytical approximations derived for large populations.

In the first section, we will explain the method used to perform the simulations and give the units used in the model. In the later sections, we compare the results of the simulations with the analytical results and discuss the roles each parameter plays. The results of the strong selection will be stated in Section 4.2. In Sections 4.3 and 4.4, we will explain how the yields y_k and trade-off parameter γ_k affect the fixation probability and the mean of the first absorption time. As big discrepancies are found, the corrected fixation probability derived in the last chapter will be compared to check whether it fits the numerical calculation better in Subsection 4.4.2.3. Lastly, in Section 4.5, the quasi-stationary distribution will be numerically calculated and compared to the Gaussian approximation derived from the analytical analysis.

4.1 Numerical Model

4.1.1 Stochastic Simulation Algorithms

The simulation approach regards the time evolution of the number of individuals of a given species as a discrete random-walk process. It can be numerically simulated at an individual level with algorithms which use a rigorously derived Monte Carlo procedure.

To perform the simulation, Gillespie's algorithm will be applied.[23] Here two questions need to be solved:

The first is, after the current state, which event occurs next? To determine this, the rates of all possible changes to the state of the model are computed, and then ordered in an array (b_1, b_2, b_3, b_4), where b_k, b_{k+2} are birth and death rates of the species type k respectively. Next, the cumulative sum of the array is taken, and the final cell contains the number R, which is the total event rate $R = \sum_{v=1}^{4} b_v$. The cumulative array is now a discrete cumulative distribution, and can be used to choose the next event by picking a random number $r_2 \sim U(0, 1)$ and choosing the μ^{th} event, such that $\sum_{v=1}^{\mu-1} \frac{b_v}{R} < r_2 < \sum_{v=1}^{\mu} \frac{b_v}{R}$.

The second question is, when does the next event occur? Since the change on the population level is discrete and governed by a Poisson process, then the time to the next event is exponentially distributed, and may be determined by a random number r_1 drawn from an exponentially distribution function with mean $\frac{1}{R}$. Thus time to next event will be advanced by $\frac{1}{R}\log(-r_1)$, given r_1 is drawn from a uniform distribution.

A diagram of the algorithms for simulating the stochastic time evolution of the chemostat model in the absence of immigration is shown in Figure 4.1.

4.1.2 Units of the Parameters in the Simulations

We analysed a continuous approximation in the limit of infinite values of V, but the numerical model is the exact discrete process with explicit individuals. A typical numerical parameters and their units are given in

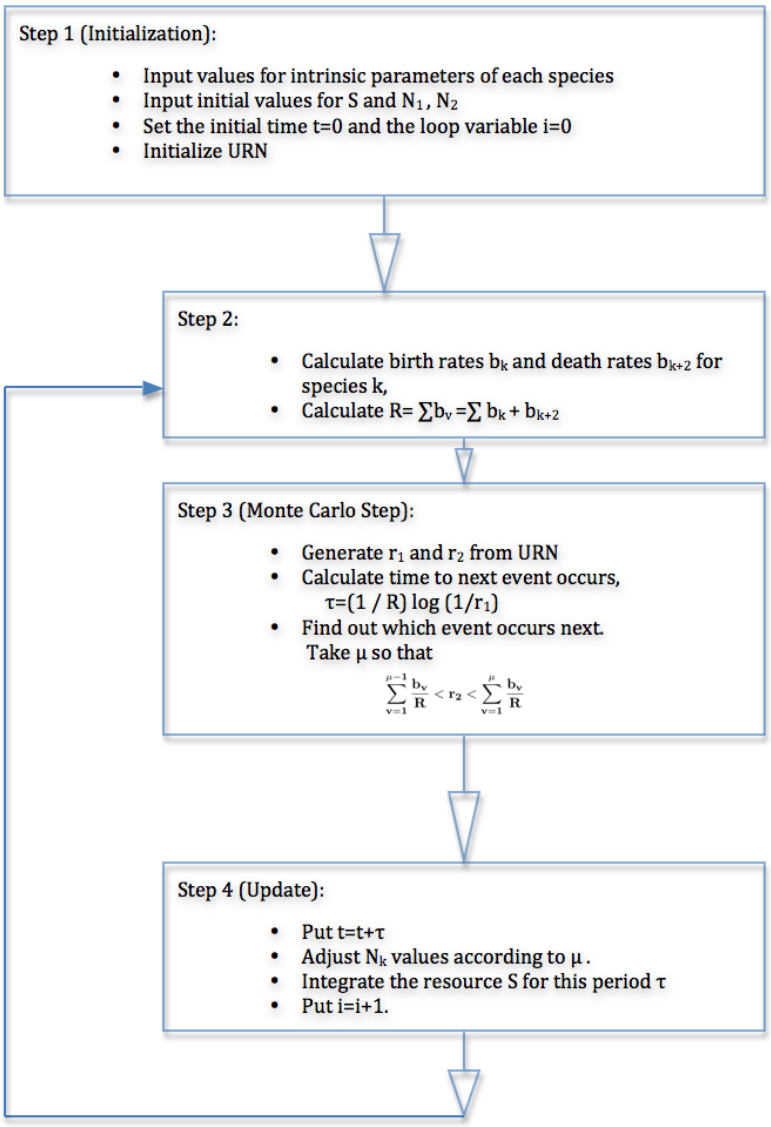

Figure 4.1: Schematic of the Stochastic Simulation Algorithm. [23]

Table 4.1 for instance.

Table 4.1: Values and Units of the Parameters in a Typical Simulations

Symbol	Name	Value and Units
V	volume	0.2 liter
y_k	yields	$5 \cdot 10^8$ cell/g
m_k	maximal growth rate	1 per hour
a_k	half saturation constant	$1 \cdot 10^{-6}$ g/liter
μ_k	intrinsic death rate	0 per hour
D	dilution rate	0.075 per hour
S^{in}	input concentration	$5 \cdot 10^{-6}$ g/liter

The choice of parameters is based on the value used in Hansen and Hubbell's experiments [26]. However, we adjust the values of the yields y_k to reach a good balance between computational time and an appropriate population size.

4.2 Strong Selection

Over finite time intervals, the stochastic dynamics fluctuate around the deterministic limit as an Ornstein-Uhlenbeck process. In this section, we will choose parameter sets for species with different deterministic fitness, and compete them numerically. The results of the simulation are then compared with the deterministic analytical strong selection results. The assumption of the deterministic dynamics means that, as the population reaches the centre manifold, the equilibrium state remains.

Effect of m_k

Firstly, we simulated the model with the parameters in Table 4.2, where the value of m_k is different between the two species.

Table 4.2: Parameters of Simulation: Different m_k

species k	S^{in}	$S(t=0)$	$N_k(t=0)$	m_k	a_k	μ_k	$V \cdot y_k$	D
1(blue)	$5 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	100	1	$1 \cdot 10^{-6}$	0	$5 \cdot 10^{7}$	0.075
2(red)	$5 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	100	1.925	$1 \cdot 10^{-6}$	0	$5 \cdot 10^{7}$	0.075

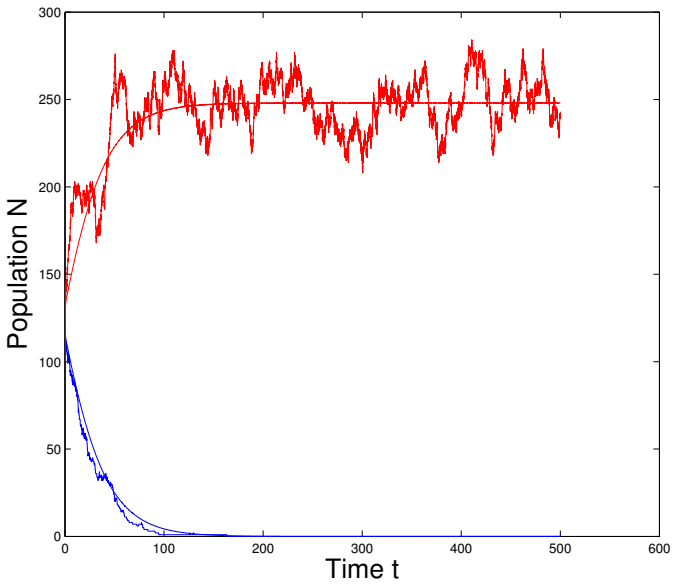

Figure 4.2: The smooth lines are deterministic result, the fluctuating lines give the stochastic numerical simulation. Red lines represent the species with higher $m = 1.925$, and blue lines are the species with $m = 1$. All other parameters are identical for the both species as in Table 4.2.

As is shown in Figure 4.2, the stochastic numerical dynamics fluctuate around the smooth deterministic lines. The red lines gives the dynamics of

the second species with the higher maximal growth rate m_k, i.e, smaller S_k^*, with a higher deterministic fitness.

Effect of a_k

Secondly, a parameter set with different a_k as in Table 4.3 will be simulated.

Table 4.3: Parameters of the Simulation: Different a_k

species k	S^{in}	$S(t=0)$	$N_k(t=0)$	m_k	a_k	μ_k	$V \cdot y_k$	D
1(blue)	$5 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	100	1	$1 \cdot 10^{-6}$	0	$5 \cdot 10^7$	0.075
2(red)	$5 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	100	1	$2 \cdot 10^{-6}$	0	$5 \cdot 10^7$	0.075

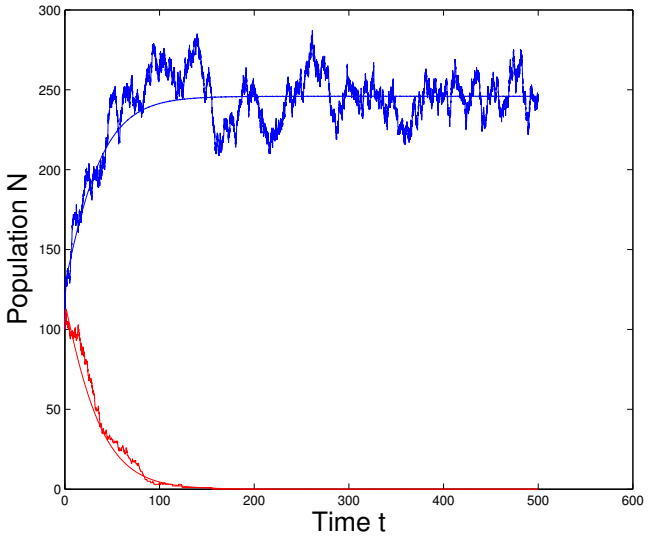

Figure 4.3: The two smooth lines are the deterministic results, the fluctuating lines are the stochastic numerical computations. Red lines represent the species with higher $a = 10^{-6}$, blue lines are the species with $a = 2 \cdot 10^{-6}$.

Figure 4.3 illustrates the convergence of the stochastic numerical results

to the smooth deterministic results, and shows that selection favours the species with the lower half-saturation constant a_k, i.e, smaller S_k^* (the blue line).

Both the numerical competitions above justify the analytical strong selection result, that species with the smaller value of $S_k^* = \frac{a_k d_k}{m_k - d_k}$ perform better. The yield y_k does not affect the fitness of the species, its effect on the weak selection will be analysed in later simulations.

Quasi-neutral Case

The numerical strong selection results for two quasi-neutral species with parameters in Table 4.4 was plotted in Figure 3.1 in the last chapter. The time phases 1 and 2 in Figure 3.1 illustrate that, once the equilibrium state is reached, two quasi-neutral species coexist for a long time and their population process fluctuate around this equilibrium value.

Table 4.4: Parameters of the Simulation: Quasi Neutral Model

species k	S^{in}	$S(t=0)$	$N_k(t=0)$	m_k	a_k	$V \cdot y_k$	D+μ
1(blue)	$2 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	100	1.925	$2 \cdot 10^{-6}$	$5 \cdot 10^{8}$	0.075
2(black)	$2 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	100	1	$1 \cdot 10^{-6}$	$5 \cdot 10^{8}$	0.075

4.3 Methodology for the Weak Selection Simulation

After the strong selection, the quasi-neutral species will coexist in the equilibrium state for a long time. From the discussion in Chapter 3, we know that eventually there will be only one species left that dominates the

whole population. In the following sections we will numerically calculate the results of the weak selection to compare with the analytical approximations. By simulating the fixation probability and the mean of the first absorption time with different parameters, we will show the effect of each parameter on the weak selection. Before giving the comparisons and discussions, it is important to clarify two points: how to calculate the fixation probability and which parameters offer the best balance between available computing resources and the accuracy of the results.

For each simulation, either the first species or the second species fixes and dominates the whole population. To calculate the fixation probability numerically, we need to run simulations under the same parameters a large number of times to estimate the probability of fixation by a certain species and the mean value of the first absorption time. Figure 4.4 shows the pseudo-code for finding the fixation probability by using the algorithms outlined in Figure 4.1, where R is the number of simulation runs.

Input: Natural number $R > 0$
Output: Estimated of fixation probability of first species

$x \leftarrow 0$ (x will count number of times the first species fixes)
for $i = 0 \rightarrow R$ **do**
  {
   Simulation in Figure 4.1
  }
  while ($N_1 \neq 0$ and $N_2 \neq 0$)
  If $N_2 = 0$ then $x++$
end for
return x/R (the estimated fixation probability)

Figure 4.4: Algorithm: Simulation to Compute Fixation Probabilities

Let the estimated fixation probabilities returned by the simulation in

Figure 4.4 be $F_R = \frac{x}{R}$ with different values of R. As seen from Figure 4.4, F_R is the mean of a R-sized sample of binomially-distributed random variables.

The standard error of the estimate for the fixation probability F_R is

$$\begin{aligned} \text{SE} &= \sqrt{\frac{\text{Sample standard deviation}^2}{\text{Number of observations}}} = \sqrt{\frac{\frac{1}{R-1}RF_R(1-F_R)}{R}} \\ &= \sqrt{\frac{F_R(1-F_R)}{R-1}} \end{aligned}$$

which decreases as the value of R increases.

However, it is not possible to run the simulation as many times as necessary to obtain a very accurate result, due to both the computational resource and time constraints. Therefore, running times and population sizes for the simulations need to be appropriately selected.

4.3.1 Choice of the Population Size and the Number of Simulation Runs Executed

4.3.1.1 Population Size

Referring to the strictly neutral model with the parameter set except y_k listed in Table 4.5, if the value of y_k is 10^9, the corresponding population size will be around 10000. It takes several hours to run one simulation with this population size. For a fixation probability which needs at least hundreds of simulation runs to compute accurately, this population size is too big. Conversely, the value of 10^7 for y_k is too small, because its corresponding population size 100 cannot ensure a good analytical approximation. There-

fore, we finally chose 8 as the order of the yields, which corresponds to a population around 1000. Most of the simulations in the following sections are run with a population around this size.

Table 4.5: Parameters for Neutral Model

species	S^{in}	$S(t=0)$	$N_k(t=0)$	m_k	a_k	μ_k	$V \cdot y_k$	D
1	$2 * 10^{-6}$	$.081 \cdot 10^{-6}$	$960p$	1	$1 \cdot 10^{-6}$	0	$5 \cdot 10^{8}$	0.075
2	$2 * 10^{-6}$	$.081 \cdot 10^{-6}$	$960(1-p)$	1	$1 \cdot 10^{-6}$	0	$5 \cdot 10^{8}$	0.075

4.3.1.2 Number of Simulation Runs Executed

It is known that in a strictly neutral model, the fixation probability equals the initial relative abundance, i.e., $f(p) = p$. Therefore, our simulations can be tested, using the neutral parameter set in Table 4.5. By comparing the numerical results to the analytical neutral line $f(p) = p$, we can adjust the number of simulation runs executed R, and hence run time, needed for an accurate result.

We ran the simulation 100, 200, 500 times respectively with $p = 0.1, 0.2 \cdots 0.9$ and plotted the results in Figure 4.5. This illustrates that all the fixation probabilities are close to the analytical neutral line $f(p) = p$, but the top figure in Figure 4.5 with $R = 100$ gives the lowest accuracy (largest standard error) compared to the results in the middle figure with $R = 200$ and bottom figure with $R = 500$. To balance the running time and accuracy, we will chose $R = 200$ for each simulation with a population over 1000, and R =500 or 1000 in simulations with population under 1000.

All simulations throughout this chapter were started from initial values

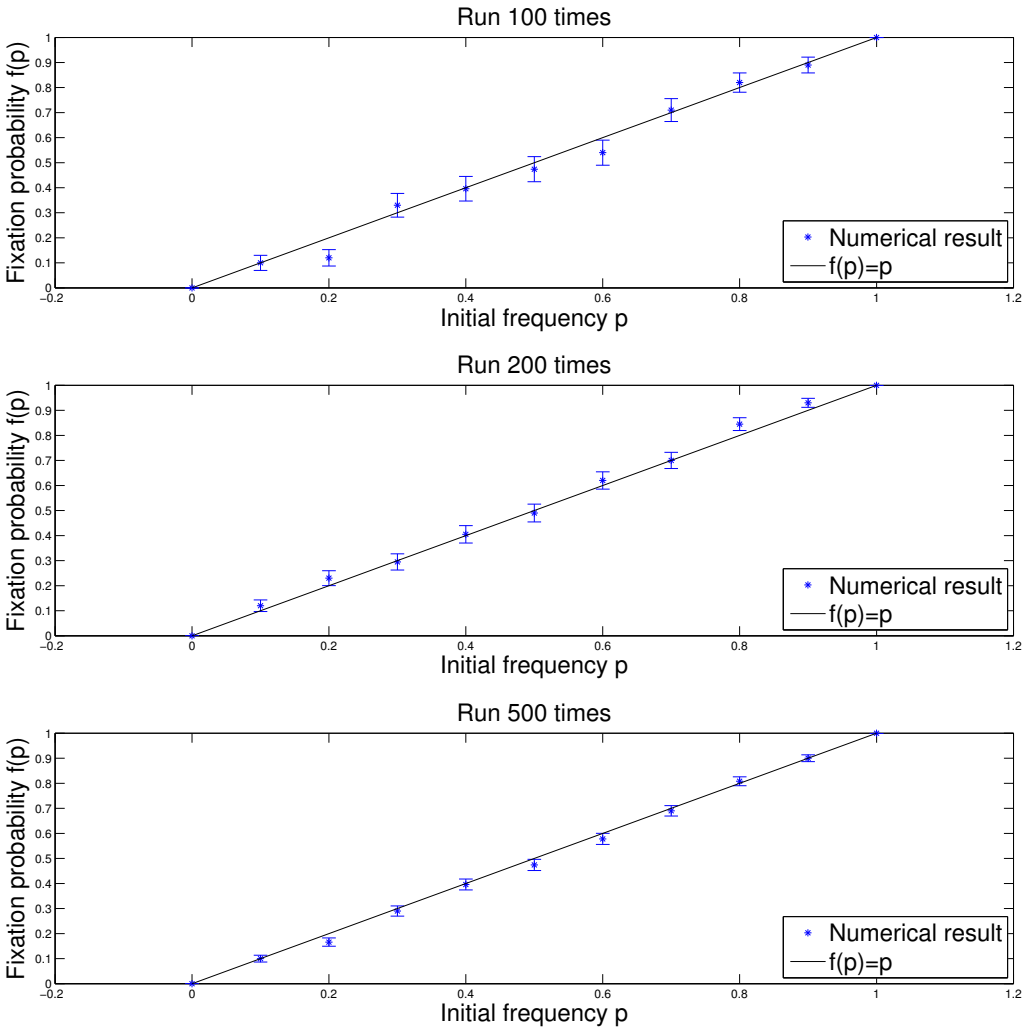

Figure 4.5: Simulation results for the fixation probability of the strictly neutral model. The dots are the simulated fixation probabilities and the lines are the true fixation probabilities, $f(p) = p$. The number of simulation runs executed for the top, middle and bottom figures are 100, 200 and 500 respectively. Error bars are standard errors.

in the equilibrium state, with the purpose of clearly understanding how the result will be affected by different initial relative abundance on the centre manifold. In the following sections, we will compare the numerical results over different parameters with the analytical results obtained in Chapter 3, to test the influence of the parameters and the accuracy of the diffusion approximation.

4.4 Effects of the Parameters on the Fixation Probability

Since the species' intrinsic death rates μ_k are assumed to be negligible compared with the dilution rate D, μ_k will be dropped from now on to simplify the comparison.

As derived in the last chapter, the fixation probability in our first approximation is

$$F(p) = \frac{\int_0^p \psi(u)du}{\int_0^1 \psi(u)du},$$

with

$$\psi(p) = \frac{(\gamma_1\kappa_1 - \gamma_2\kappa_2)p + \gamma_2\kappa_2}{[(\kappa_1 - \kappa_2)p + \kappa_2]^3} \left\{ \frac{(\gamma_1^2 - \gamma_2^2)p + \gamma_2^2}{(\kappa_1 - \kappa_2)p + \kappa_2} \right\}^{\frac{(\gamma_1\kappa_2 - \gamma_2\kappa_1)(\gamma_1+\gamma_2)}{\gamma_2^2\kappa_1 - \gamma_1^2\kappa_2}}.$$

In Subsections 4.4.1 and 4.4.2, these analytical predictions for the fixation probability will be compared with the numerical results to determine the effects of the yields y_k and γ_k respectively. Since large discrepancies will appear, in Subsection 4.4.2.3, we will show that the corrected fixation probability derived in Subsection 3.7.2 matches the numerical results better.

4.4.1 Effects of the Yields y_k

4.4.1.1 Effects of the Yields y_k, While γ_k Are the Same

To remove the impact of all other parameters, here we assume that the parameter γ_k is identical to both species, namely $\frac{\gamma_1}{\gamma_2} = 1$. Using the analytical results in Subsection 3.7.1.3, we have

$$F(p) = p,$$

regardless of the value of $\frac{y_1}{y_2}$. That is, under the situation where the birth rate $b_k = \frac{m_k S}{a_k + S}$ and its sensitivity to the resource at the equilibrium state $\gamma_k = \frac{\mathrm{d}b_k}{\mathrm{d}S}|_{S^*}$ are identical to both species at the equilibrium state, the fixation probability is only dependent on its initial frequency, as in the strictly neutral model.

To numerically test the negligible influence of the yields on the fixation probability under the condition of identical γ_k, we expand the conditions slightly. Let the value of γ_1 and γ_2 be close but not exactly identical. Other values of the parameters are listed in Table 4.6 where $\frac{\gamma_1}{\gamma_2} = 0.9625$. The numerical results of the fixation probability are plotted together with the analytical predictions in Figure 4.6.

Table 4.6: The Parameter set to test the effect of the yields

species k	S^{in}	$S(t=0)$	m_k	a_k	μ_k	D
1	$2 \cdot 10^{-6}$	$.081 \cdot 10^{-6}$	1	$1 \cdot 10^{-6}$	0	0.075
2	$2 \cdot 10^{-6}$	$.081 \cdot 10^{-6}$	1.925	$2 \cdot 10^{-6}$	0	0.075

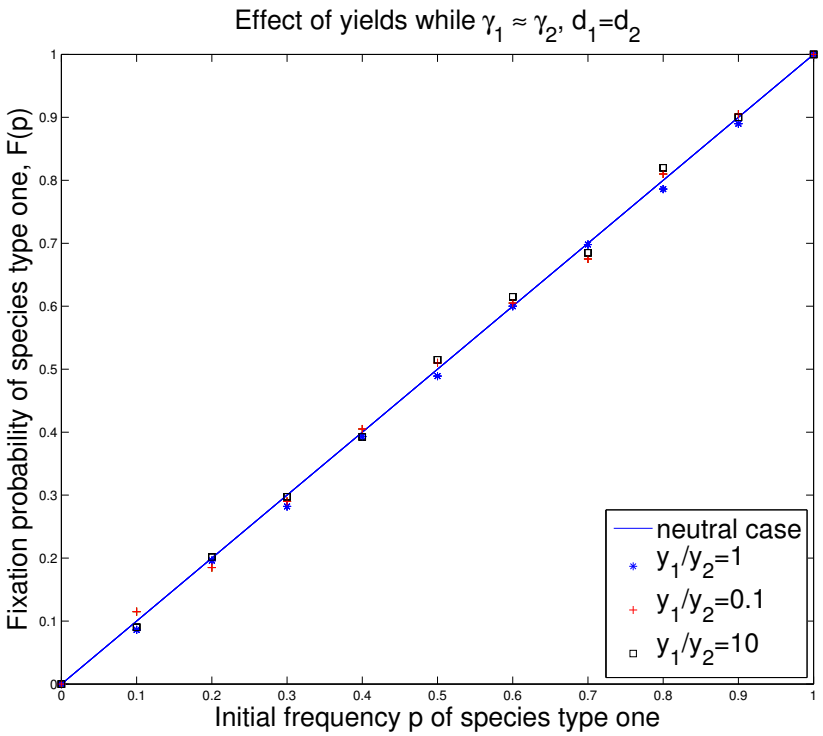

Figure 4.6: The effect of the yields on the fixation probability when $\frac{\gamma_1}{\gamma_2} = 0.9625$ and $\frac{d_1}{d_2} = 1$. The dots are the results of the numerical calculation, and the lines are the analytical results. The results for different ratios of the yields $\frac{y_1}{y_2} = 1, 0.1, 10$ are plotted in the colours of blue, red and black respectively.

Figure 4.6 illustrates that the numerical fixation probabilities are all located around the the analytical neutral lines no matter how the yields changes. This close match between the numerical and analytical results demonstrates that if the trade-off parameters γ_k are close to each other for the quasi-neutral species, then a large variation in yields do not affect the fixation probability result greatly.

It was shown in Subsection 3.8.1 that the species in the quasi-stationary phase is Gaussian distributed with the mean and variance given in Eqs.(3.66)-(3.69). If γ_1=γ_2, these two Gaussian processes have the same magnitude of relative fluctuations equal to :

$$\frac{\delta_1^2}{\mu_1} = 1 + \frac{D}{\gamma_1(S^{\text{in}} - S^*)} + \frac{D}{D + \gamma_1(S^{\text{in}} - S^*)} = \frac{\delta_2^2}{\mu_2}.$$

As discussed in Subsection 2.4.1, if $\gamma_1 = \gamma_2$, then all other trade-off parameters are identical.

Therefore, if the quasi-neutral species has the same trade-off parameters, they will be ecologically neutral in the weak selection no matter how different their yields are.

4.4.1.2 Effect of the Yields y_k, While γ_k Are Different

However, this is only the case when $\gamma_1 = \gamma_2$. If the γ_k or d_k are very different from each other then the yields do impact the fixation probability. This is because, when $\gamma_1 \neq \gamma_2$, the sign of the drift coefficient $b(q)$ is dependent on all the parameters p, γ_k and y_k. In Figure 4.7, analytical results for the fixation probability are shown for decreasing and increasing $\frac{y_1}{y_2}$ with $\gamma_1 > \gamma_2$ and $\gamma_1 < \gamma_2$. These results are complex and the derivation from neutrality does not depend on any single parameter.

4.4.2 Effects of the Parameter γ_k

The sensitivity of the birth rate to the resource at the equilibrium state, γ_k, is an important parameter in the model. In the deterministic analy-

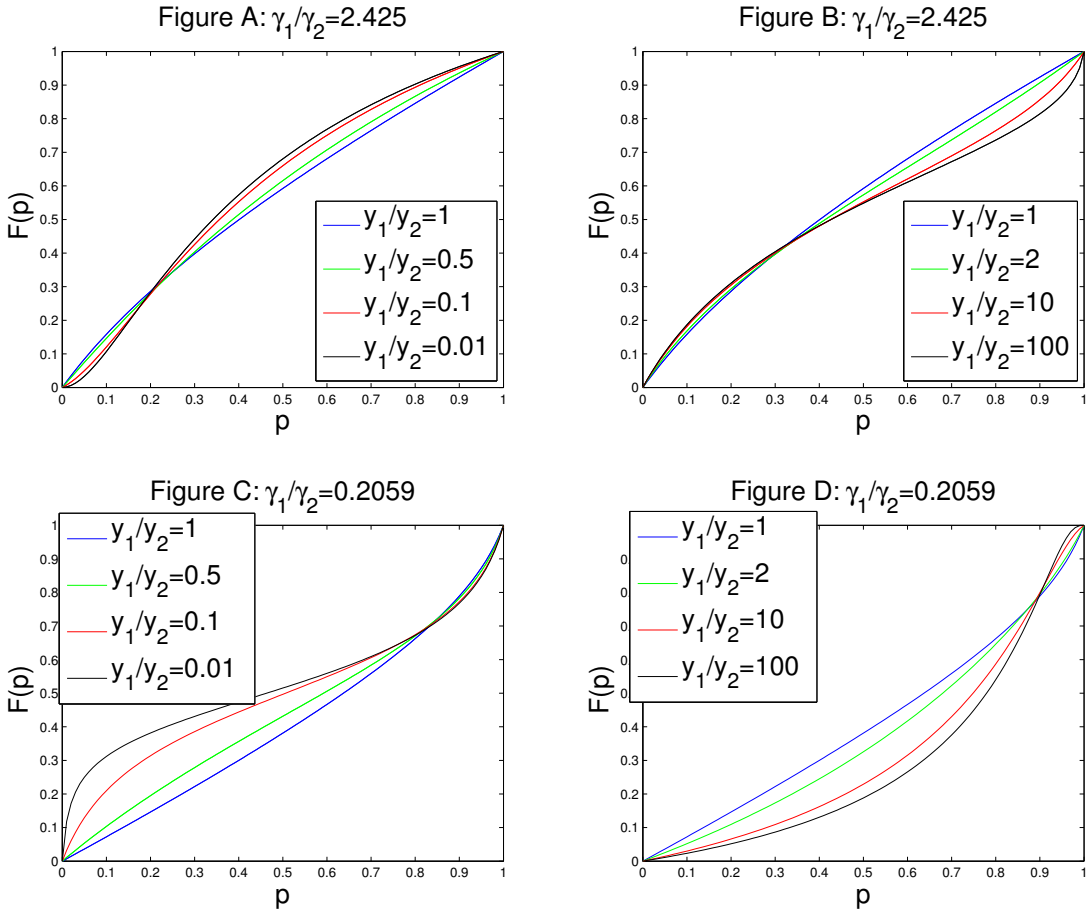

Figure 4.7: The effect of the yields on the analytical fixation probability while $\gamma_1 \neq \gamma_2$. In Figure A and B, $m_1 = 1,\ m_2 = 0.12125,\ a_1 = 1{\cdot}10^{-6},\ a_2 = 5 \cdot 10^{-8}$. In Figure C and D, $m_1 = 0.0935,\ m_2 = 1.925,\ a_1 = 2 \cdot 10^{-8},\ a_2 = 2 \cdot 10^{-6}$. The dilution rate $D = 0.075$.

sis, the species with larger γ_k grows more quickly before saturation and achieves a larger relative abundance when it reaches the centre manifold. In this subsection, we will analyse the effects of varying γ_k on the fixation probability $F(p)$. The yields y_k will be kept fixed at the same value for the

two quasi-neutral species throughout this section.

4.4.2.1 Analytical Results

Given that $y_1 = y_2, d_1 = d_2$, using the results in Subsection 3.7.1.2, the analytical fixation probability using our first approximation is:

$$F(p) = p + \frac{\frac{\gamma_1}{\gamma_2}[\log((\{\frac{\gamma_1}{\gamma_2}^2\} - 1)p + 1) - 2p\log\frac{\gamma_1}{\gamma_2}]}{\frac{\gamma_1}{\gamma_2}^2 - 1 + 2\frac{\gamma_1}{\gamma_2}\log\frac{\gamma_1}{\gamma_2}}.$$

This shows that the species with large γ_k has an increased fixation probability.

This advantage can be explained by the difference between their quasi-stationary distributions in the quasi stationary phase. Given the same yield rate, the quasi-neutral species have the same mean abundance,

$$x_1^* = (S^{\text{in}} - S^*)y_1 = (S^{\text{in}} - S^*)y_2 = x_2^*.$$

However, the species with larger γ has a smaller variance,

$$\gamma_1 > \gamma_2$$
$$var_1 = \frac{1}{V}\{D(\frac{x_1*}{D} + \frac{y_1}{\gamma_1}) + \frac{D}{\frac{D}{x_1^*} + \frac{\gamma_1}{y_1}}\} < \frac{1}{V}\{D(\frac{x_2^*}{D} + \frac{y_2}{\gamma_2}) + \frac{D}{\frac{D}{x_2^*} + \frac{\gamma_2}{y_2}}\} = var_2.$$

Therefore, the species with a birth rate that is more sensitive to the resource concentration at the equilibrium state is more likely to dominate the population in the weak selection regime, and following fixation has a more stable population in the quasi stationary phase.

4.4.2.2 Numerical Comparison

In order to compare these predictions with the numerical results, we fix S^*, the values of the yields and the species death rates at constant values, that $S^* = 0.081 \cdot 10^{-6}$, $S^{\text{in}} = 2 \cdot 10^{-6}$ and $y_1 = y_2 = 5 \cdot 10^8$, $d_1 = d_2 = 0.075$. Due to the yields being the same, the total expected population size will not change with the different initial relative abundance p of the first species, that is,

$$\Pi_1 + \Pi_2 = (S^{\text{in}} - S^*)y = 960.$$

In Figure 4.8, we simulate numerical models with different ratios of γ_k ($\frac{\gamma_1}{\gamma_2} = 0.9625,\ 1.3,\ 2.425$) and compare them with the analytical predictions. In both case we observe an advantage to the species with larger γ_k in the weak selection regime.

4.4.2.3 Better Fit Using the Corrected Fixation Probability

In Figure 4.8, it is apparent that there are discrepancies between the numerical results and the analytical approximations. By increasing the population size from 1000 to 10000, the running time for one simulation increases greatly. However there is no even tiny apparent improvement of shrinking the discrepancies. It suggests that the analytical approximation is not sufficiently accurate. Such discrepancies are caused by the absence of the higher order term of Taylor expansion when deriving the derivatives of the projection map. Based on the above fact, in this subsection, we will use the results for the corrected fixation probability in Subsection 3.7.2 to compare with the numerical results.

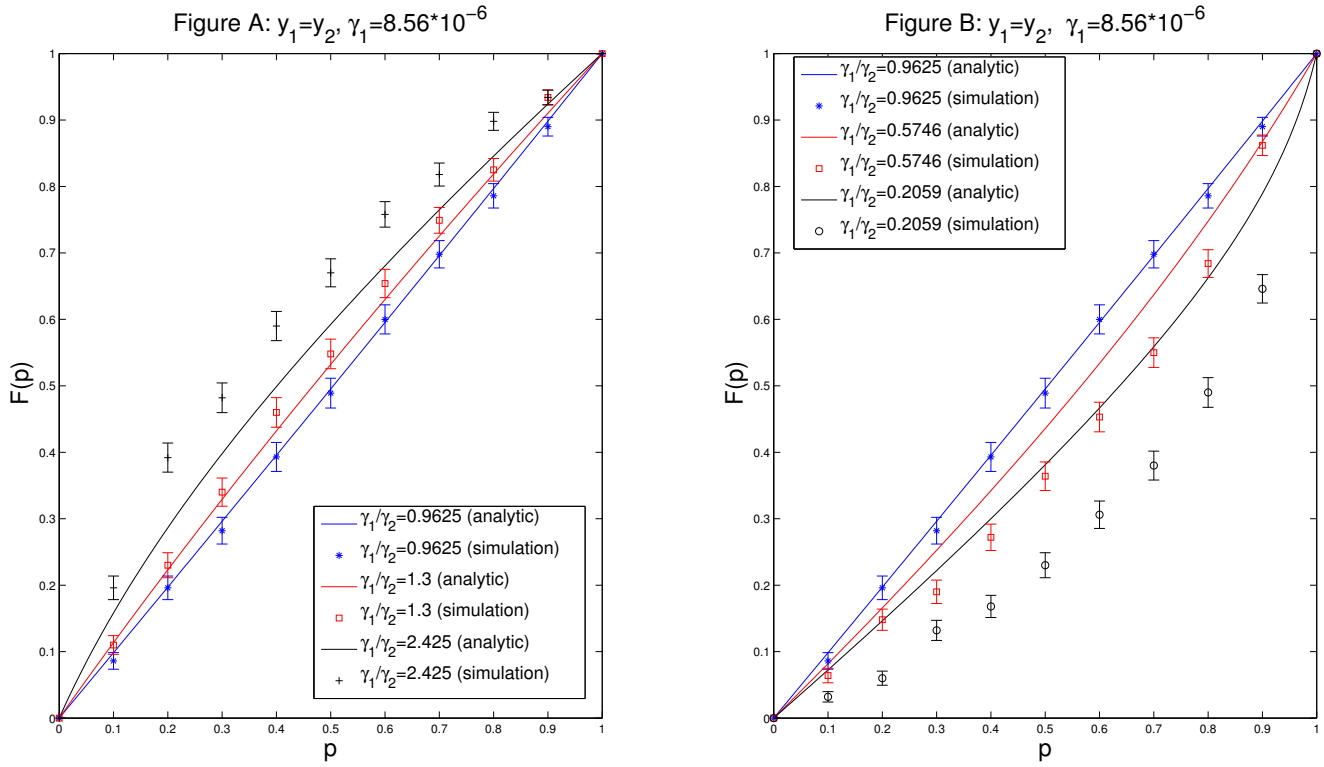

Figure 4.8: The effects of γ_k on the fixation probability: a comparison of the numerical results to the analytical prediction. In Figure A: $a_1 = 1 \cdot 10^{-6}$, $m_1 = 1$. $a_2 = 2 \cdot 10^{-6}$, $0.2 \cdot 10^{-6}$, $5 \cdot 10^{-8}$, and $m_2 = 1.925$, 0.26, 0.12125 in blue, red and black lines respectively. In Figure B: $a_2 = 2 \cdot 10^{-6}$, $m_2 = 1.925$. $a_1 = 1 \cdot 10^{-6}$, $1 \cdot 10^{-7}$, $2 \cdot 10^{-8}$, and $m_1 = 1$, 0.1675, 0.0935 in blue, red and black lines respectively. $y_1 = y_2 = 5 \cdot 10^{-8}$ in all calculations. Error bars are standard errors.

Given that $y_1 = y_2$, $d_1 = d_2 = D$, with the expressions in Eqs.(3.59)

and (3.62), the fixation probability is expressed as,

$$\psi_1(p) = \left(\frac{(\frac{\gamma_1}{\gamma_2} - 1)p + 1}{(\frac{\gamma_1}{\gamma_2}^2 - 1)p + 1}\right)^{1 - \frac{D^2}{S^*(\gamma_1 - \gamma_2)}(\frac{1}{m_1} - \frac{1}{m_2})}$$
$$F(p) = \frac{\int_0^p \psi_1(u)\mathrm{d}u}{\int_0^1 \psi_1(u)\mathrm{d}u}.$$

Using the same parameters as in Subsection 4.4.2.2 and Figure 4.8, the corrected analytical results for the fixation probability are plotted in Figure 4.9 together with the numerical results. A very good match between the analytical and numerical fixation probabilities is obtained.

It is explained in Subsection 3.7.2 that the two sets of derivatives of the projection map will result in a difference in drift coefficient $b(p)$ of the diffusion approximation of the relative abundance, but not the diffusion coefficient $a(p)$. The reason why the corrected approximation makes such a large improvement is due to $b(p)$ having a significant impact on the fixation probability. However, in the calculation of the mean of the first absorption time, the coefficient $a(p)$ has greater impact. In that case the approximation derived from the first set of derivatives of the projection map is sufficient to obtain a good match to the numerical calculations.

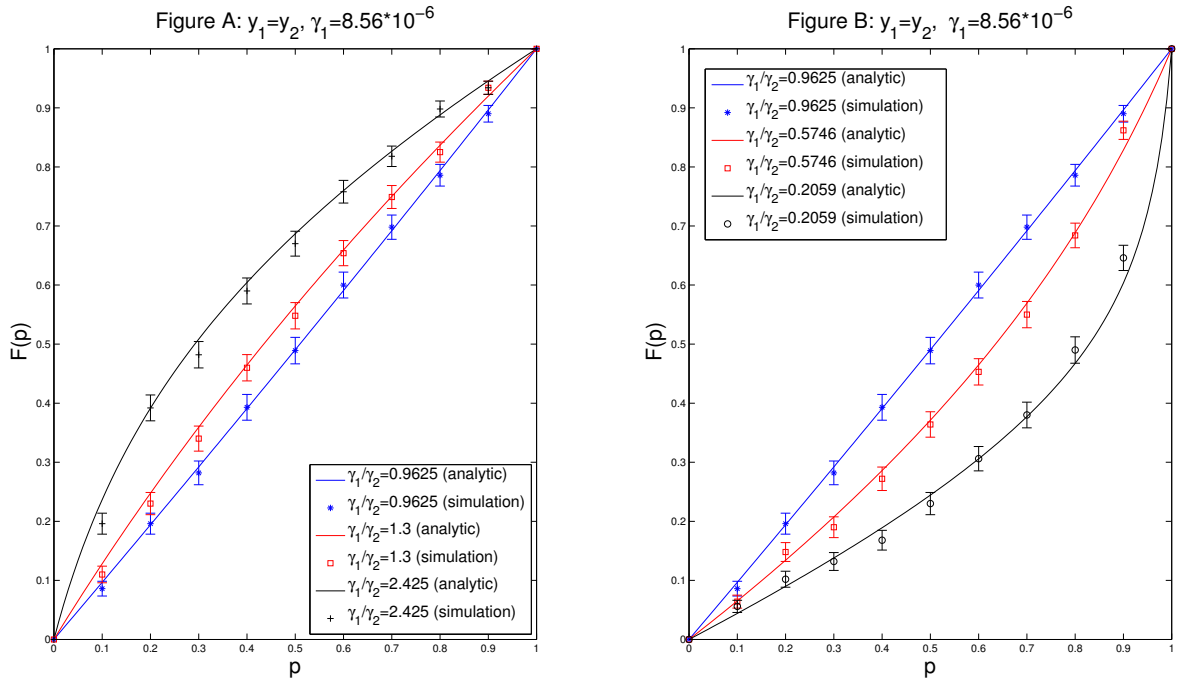

Figure 4.9: The effects of γ_k on fixation probability using the corrected analytical approximations. In Figure A: $a_1 = 1 \cdot 10^{-6}$, $m_1 = 1$. $a_2 = 2 \cdot 10^{-6},\ 0.2 \cdot 10^{-6},\ 5 \cdot 10^{-8}$, and $m_2 = 1.925,\ 0.26,\ 0.12125$ in blue, red and black lines respectively. In Figure B: $a_2 = 2 \cdot 10^{-6}$, $m_2 = 1.925$. a_1 equal $1 \cdot 10^{-6},\ 1 \cdot 10^{-7},\ 2 \cdot 10^{-8}$, and m_1 equal $1,\ 0.1675,\ 0.0935$ in blue, red and black lines respectively. $y_1 = y_2 = 5 \cdot 10^{-8}$ in all calculations. Error bars are standard errors.

4.5 Effects of the Parameters on the Mean of First Absorption Time

The analytical result for the mean of the first absorption time $ET(p)$ is calculated in Section 3.7.2,

$$ET(p) = 2\int_0^1 \psi(x)\mathrm{d}x \cdot \Big\{(1-F(p))\int_0^p \frac{F(x)}{a^2(x)\psi(x)}\mathrm{d}x + F(p)\int_p^1 \frac{1-F(x)}{a^2(x)\psi(x)}\mathrm{d}x\Big\}.$$

In this section, we will analyse how the parameters affect these results.

4.5.1 Effects of γ_k

To analyse the effect of γ_k on the mean of the first absorption time, we use the same parameter sets as in Subsection 4.4.2.2, where $S^* = 0.081 \cdot 10^{-6}$, $S^{\text{in}} = 2 \cdot 10^{-6}$, $y_1 = y_2 = 5 \cdot 10^8$, and $d_1 = d_2 = D = 0.075$. Due to the yields being identical, the total expected population size remains constant,

$$\Pi_1 + \Pi_2 = (S^{\text{in}} - S^*)y = 960,$$

for different initial relative abundance value p. The analytical and numerical results for the means of the first absorption times with different γ_k are plotted in Figure 4.10.

As Figure 4.10 shows, the analytical results derived by the diffusion approximation match well with the numerical results even for small populations. The means of the first absorption time do not change greatly with the different $\frac{\gamma_1}{\gamma_2}$ ratio. Therefore, we suggest that the yields-dependent population size has a larger impact on the mean of the first absorption time than the parameter γ_k. To verify this, in the next subsection, we will change y_k and population size to see how the mean of the first absorption time varies.

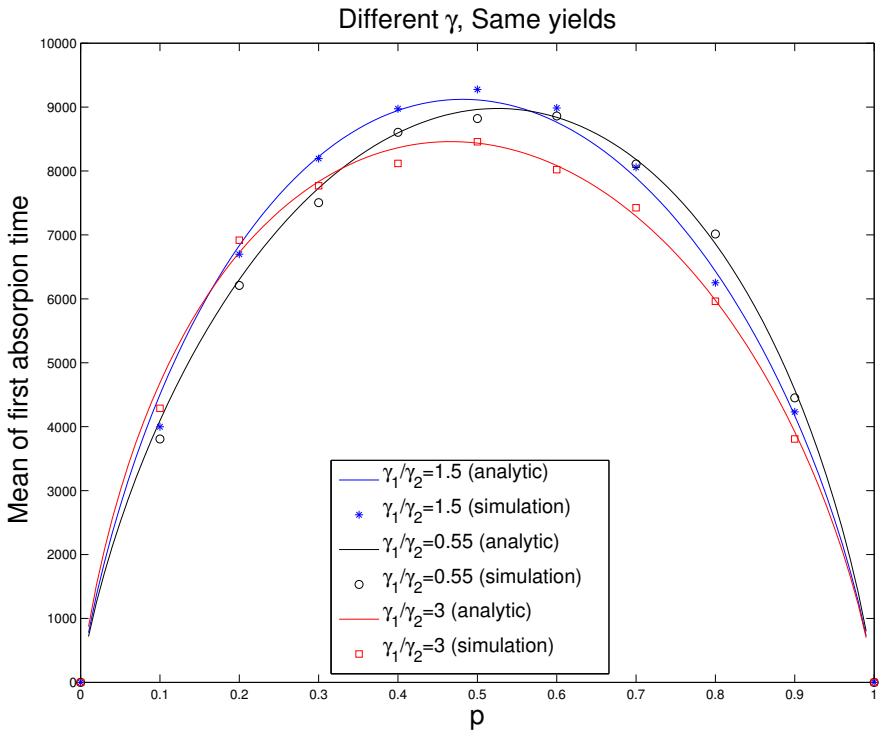

Figure 4.10: The mean of first absorption time with different γ_k. The parameters of the first species are kept constant as $a_1 = 1 \cdot 10^{-4}$, $m_1 = 0.15$. For the second species in the blue, black and red lines, $a_2 = 5 \cdot 10^{-4}$, $1.1 \cdot 10^{-3}$, $2 \cdot 10^{-5}$, $m_2 = 0.1125,\ 0.9,\ 0.09$ respectively. $y_1 = y_2 = 10^7$,$D = 0.075$.

4.5.2 Effects of the Yields y_k

As derived in Subsection 3.7.2, the mean of the first absorption time in the strictly neutral case is

$$ET(p) = -y(S^{\text{in}} - S^*)\Big(p \log p + (1-p)\log(1-p)\Big),$$

where $ET(p)$ is proportional to the total expected population size $\Theta = y(S^{\text{in}} - S^*)$.

In the quasi-neutral model, if we set the γ_k to be identical, the expected population size varies dramatically with the initial relative abundance p,

$$\Theta = \Pi_1 + \Pi_2 = \frac{(S^{\text{in}} - S^*)D}{\kappa_1 p + \kappa_2(1-p)}$$

Using the results in Subsection 3.7.2, we have the mean of the first absorption time,

$$\begin{aligned} ET(p) &= \frac{1}{D}\Big((1-p)\int_0^p \frac{\Theta}{1-x}\mathrm{d}x + p\int_p^1 \frac{\Theta}{x}\mathrm{d}x\Big) \\ &= (\frac{1-p}{\kappa_1} + \frac{p}{\kappa_2})\log\Big((\frac{\kappa_1}{\kappa_2} - 1)p + 1\Big) - \frac{1-p}{\kappa_1}\log(1-p) - \frac{p}{\kappa_2}\log p \end{aligned}$$

where, $\kappa_k = \frac{D}{y_k}$. Plotting this analytical result in Figure 4.11, we can determine that the yield y_k plays a major role in the mean of the first absorption time. Given an initial value p, increasing either of the yields y_k will increase the mean of the first absorption time. Increasing the value of y_2 gives an positive skew[1], to the blue and purple line. It means that the large initial relative abundance of second species will delay the first absorption time. Decreasing it, the skew is then negative[2] in the red line, and tell us that increase the initial relative abundance of the first species can help us extend the time of the coexistence phase.

[1]Positive skew indicates that the tail on the right side is longer or fatter than the left side

[2]Negative skew indicates that the tail on the left side is longer or fatter than the right side

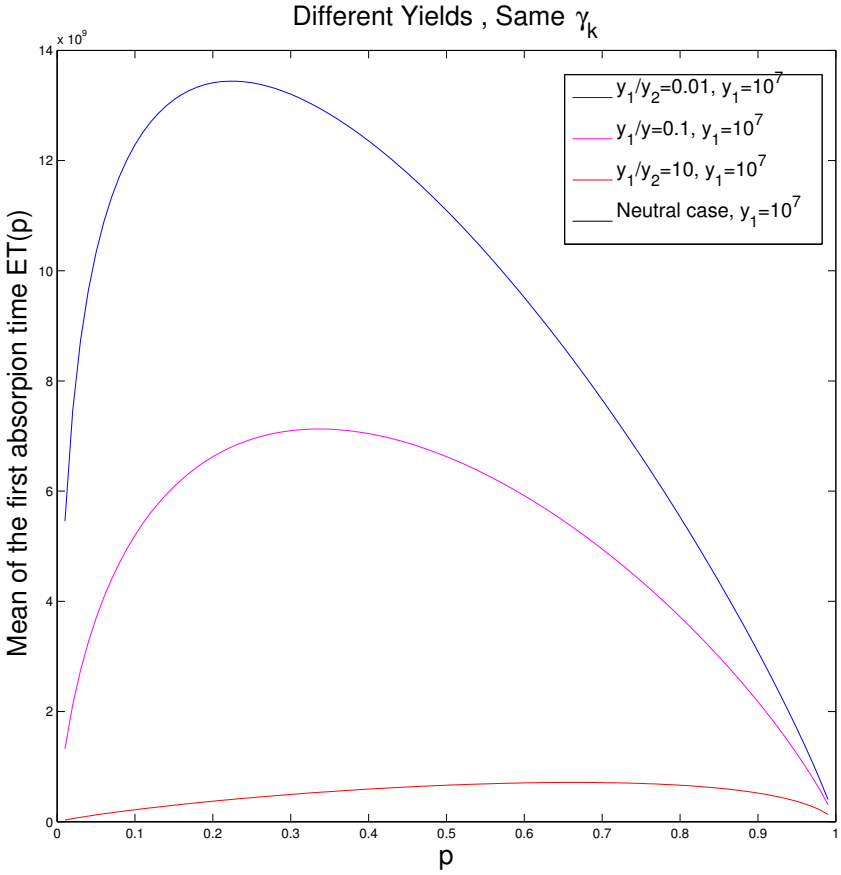

Figure 4.11: The mean of first absorption time with different yields y_k, and $a_1 = a_2 = 1 \cdot 10^{-4}$, $m_1 = m_2 = 0.15$, $S^* = 1 \cdot 10^{-4}$, $S^{\text{in}} = 2 \cdot 10^{-4}$.

4.6 Numerical Calculation of the Quasi-stationary Distribution

The theoretical assumption underlying the numerical calculation of the quasi-stationary distribution is that the progress N_k in the quasi stationary

phase is ergodic[3]. We assume a set of parameters for the two quasi-neutral species and their chemostat environment in Table 4.7.

Table 4.7: Parameter set for quasi-stationary distribution calculation

species	S^{in}	$S(t=0)$	$N(t=0)$	m_k	a_k	μ_k	$y_k * V$	D
1	$1.081 \cdot 10^{-6}$	$.081 \cdot 10^{-6}$	4000	1.95	$2 \cdot 10^{-6}$	0	$1 \cdot 10^{9}$	0.075
2	$1.081 \cdot 10^{-6}$	$.081 \cdot 10^{-6}$	16000	1	$1 \cdot 10^{-6}$	0	$1 \cdot 10^{9}$	0.075

According to the analytical quasi-stationary result in Eqs.(3.64) and (3.65), the mean and variance of the population of the first species is

$$\mu_1 = V x_1^* = V \frac{y_1 (S^{\text{in}} - S^*) D}{d_1}$$
$$\sigma_1^2 = V \Big(\frac{y_1 (D + \frac{\gamma_1}{y_1} x_1^*)}{\gamma_1} + \frac{x_1^* d_1}{D + \frac{\gamma_1}{y_1} x_1^*} \Big).$$

Given the value of the parameters in Table 4.7, the above equations yield a mean of 1000, and a standard deviation of 34.1796.

For the simulation, we will use the history up to time t (which is large enough) from an appropriate starting time t_0 in the quasi stationary phase to estimate the $f_{q,1}(N)$ with the help of the normal fit command (a matlab command). The numerical results in Figure 4.12 show that the process visits every state according to a Gaussian distribution with mean 1002 and standard deviation 34.29 which matches the analytical results perfectly.

[3]In a Markov chain, a state is said to be ergodic if it is aperiodic and non-absorbing. If all states in a Markov chain are ergodic, then the chain is said to be ergodic.

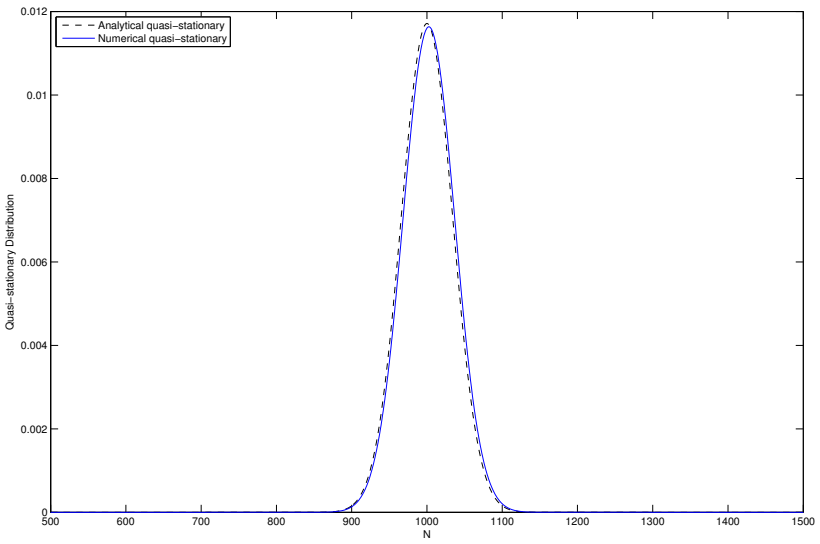

Figure 4.12: The quasi-stationary distribution, $f_{q,1}(N)$. The solid line is numerical calculation with mean 1002 and standard deviation 34.29, compared to the analytical approximation (dot line) with mean 1000 and standard deviation 34.18

4.7 Summary

This chapter compares the analytical diffusion approximations with the numerical results, and discusses how the parameters (yields y_k, sensitivity of the birth rate to the resource at the equilibrium state γ_k) affect the results of the strong selection, the fixation probability, the mean of the absorption time and the quasi-stationary distribution.

Firstly we observe, from comparing the numerical and analytical results, that if the first set of the derivatives of the projection map in Subsection 3.5.3 is used to derive the fixation probability, there is no good fit between the analytical approximation in the large population limit and numerical

results for small population. The large discrepancies between numerical and analytical results do not shrink if the simulation population size increases, and are caused by the absence of the higher order term of the Taylor expansion of the deterministic trajectory. By using the corrected results for the fixation probability, a good approximation between the numerical calculation and analytical calculation are achieved.

Secondly we conclude, that the yields y_k and the sensitivity of the birth rate to the resource at the equilibrium state, γ_k, play different roles in impacting the strong and weak selection. In the strong selection, the yields has no impact on the results, and the species with larger γ_k has the advantage of growing quickly and occupying a larger relative abundance in the long time coexistence phase. In the weak selection, it is hard to say which single parameter will be an advantage or disadvantage for the species to compete. If the γ_k are identical for both quasi-neutral species, the yields will not influence the results of weak selection, and the fixation probability is only dependent on the initial relative abundance p. Conversely, under the situation that the yields y_k are identical, there is a selective advantage for the species with larger γ_k. It will obtain a larger relative abundance during the strong selection, and have an increased fixation probability in the weak selection and increased chance to dominate the population. However, larger yields which lead to larger population will delay the time to fixation.

In all the above chapters, most results we found useful in the chemostat model with immigration absent were derived. Its deterministic model was built in Chapter 2, the stochastic model was developed in Chapter 3, and the comparisons of these two models are discussed in this chapter. In the

following chapter, we will add dispersal (immigration) to the present model to prevent fixation and maintain a stationary distribution, then determine how the disperse controls the diversity of the local communities.

Chapter 5

Quasi-Neutral Model with Immigration

In the last three chapters, the chemostat models considered ignored immigration, i.e., the influx comprised only resource. However, in the real world, most microbial communities are not isolated. A community in the absence of immigration will quickly lose diversity due to the deterministic strong selection and more slowly through stochastic weak selection. Therefore, to maintain a complex community with a stationary total diversity, immigration will be necessary. In this chapter we incorporate a steady rate of immigration of individuals into our local community from a fixed distribution in a presumed meta-community. No matter how low the rate of immigration this ensures that no species will go extinct in the system.

5.1 Introduction

Assume a small rate of population influx for each species into the local chemostat community, denoted $R_k (k = 1, 2)$, from a meta-community. Given large V, we define the rate of immigration density $r_k = \frac{R_k}{V}$ to be small, with definition [1] [2]

$$r_k = O(1/V)$$
$$R_k = o(V).$$

Namely, $r_k \to 0$, $R_k = V r_k < \infty$.

Define q_k to be the relative abundance of the species k in the meta-community. Since the immigration population enters the local community through the influx (with dilution rate D), the immigration rate of the kth species to the local community R_k, is proportional to the relative abundance of the kth species in the meta-community q_k and the dilution D:

$$\frac{R_k}{V} = r_k \propto q_k D. \tag{5.1}$$

In the deterministic perspective, this is a constant rate of immigration. In the stochastic perspective, the immigration occurs according to a Poisson distribution with jump intensity r_k.

To understand this new dynamics, firstly the deterministic model will be constructed in Section 5.2 and be approximated by the deterministic model

[1] If $f(N) = O(g(N))$, then $\lim_{N\to\infty} \sup |\frac{f(N)}{g(N)}| < \infty$
[2] If $f(N) = o(g(N))$, then $\lim_{N\to\infty} |\frac{f(N)}{g(N)}| = 0$

in the absence of immigration since the immigration rate is small. Then we will have the same results for the strong selection and the same centre manifold as in the model with immigration absent.

Later in Section 5.3, the stochastic model will be developed to derive the diffusion approximation at the long time scale. The immigration will drive the long term drift and balance extinction to maintain the diversity of the local community. With the help of the forward Kolmogorov equation, the stationary distribution of the relative abundance will be calculated in Section 5.4. In Section 5.5, the impact of the parameters on the stationary distribution is analysed. Section 5.6 is the comparison section, where the stationary distribution will be numerically calculated to determine how well this fits the analytical results. In the last section, we will compare our quasi-neutral model with Hubbell's neutral model. By analysing how the models differ, we reveal the roles the drift, immigration and species difference play in shaping the structure of the local community.

5.2 Deterministic Approximation

By the law of large numbers, the deterministic ODE system is the limit of the stochastic system over compact time intervals. In this section, the deterministic dynamics of the model with immigration will be clearly presented.

Define the deterministic dynamics of the resource concentration and

species density $(\bar{S}(t), \bar{x}_k(t))$ as the solution of the ODE system:

$$\begin{aligned}
\frac{\partial \bar{S}}{\partial t} &= \Big(S^{\text{in}} - \bar{S}(t)\Big)D - \sum_{k=1}^{2} \frac{\bar{x}_k(t)}{y_k} \frac{m_k \bar{S}(t)}{a_k + \bar{S}(t)}, \\
\frac{\partial \bar{x}_k}{\partial t} &= r_k + \frac{m_k \bar{S}(t)}{a_k + \bar{S}(t)} \bar{x}_k(t) - \Big(D + \mu_k\Big)\bar{x}_k(t), \\
\bar{S}(0) &= S_0 > 0, \\
\bar{x}_k(0) &= x_{k,0} > 0, \qquad\qquad k = 1, 2, \ldots, n.
\end{aligned}$$

Given $V \to \infty$, the small immigration density rate is negligible, i.e, $r_k \to 0$. With this limit, this ODE system can be approximated by the ODE system of the deterministic model in the absence of immigration derived in Chapter 2. Therefore, the assumption of low immigration rate $r_k \to 0$ ensures the same results for strong selection, only the quasi-neutral species at the largest fitness $\frac{1}{S^*}$ survive, and the equilibrium states on the same centre manifold, where

$$(S^{\text{in}} - S^*)D = \sum_{k}^{2} \frac{d_k}{y_k} x_k^* \tag{5.2}$$

$$S^* = \frac{a_k d_k}{m_k - d_k}. \tag{5.3}$$

The deterministic equilibrium state of the species density x_k^* are dependent on the initial values.

With the results for niche assembly unchanged, in the following sections, we determine how the limited dispersal will shape the diversity in local communities together with stochastic drift.

5.3 Stochastic Analytical Model

In the rest of this chapter, we will analysis the stochastic dynamics of the chemostat model with immigration present. Using the law of large numbers, the stochastic dynamics may be converged to the deterministic dynamics over compact time intervals, where the impact of the rare immigration density rate is negligible. The diffusion approximation over compact time intervals will be the same as derived in Chapter 3. However, over an infinite time interval, the boundaries are no longer absorbing given the presence of immigration. Although the quasi-neutral species continue to coexist, the diversity over an infinite time interval is different from that over compact time intervals. In this section, we will derive the diffusion approximations over the compact time intervals and infinite time intervals respectively, and explore the final stationary distribution of the relative abundance of the first species, $f(p)$.

5.3.1 Stochastic Model and Diffusion Approximation Over Compact Time Intervals

Compared to the stochastic model in the absence of immigration, the birth events of the species are no longer intrinsic only, as the immigration events need to be accounted for. Therefore the transition probabilities of

the population $\hat{N}_k(t)$ are different from those in Section 3.2.1,

$$P\{\hat{N}_k(t+\triangle t)=\hat{N}_k(t)+1\}=\Big(\frac{m_kS(t)}{a_k+S(t)}\hat{N}_k(t)+\underline{\underline{R_k}}\Big)\triangle t+o_1(\triangle t) \quad (5.4)$$
$$=\beta_{k,1}\Big(\hat{S}(t),\hat{x}(t)\Big)V\triangle t+o_1(\triangle t),$$
$$P\{\hat{N}_k(t+\triangle t)=\hat{N}_k(t)-1\}=(D+\mu_k)\hat{N}_k(t)\triangle t+o_2(\triangle t) \quad (5.5)$$
$$=\beta_{k,-1}\Big(\hat{S}(t),\hat{x}(t)\Big)V\triangle t+o_2(\triangle t),$$

where $\beta_{k,l}$ is the jump intensity defined as,

$$\beta_{k,1}\Big(\hat{S}(t),\hat{x}(t)\Big)=b_k(t)\hat{x}_k(t)+r_k=\frac{m_k\hat{S}(t)}{a_k+\hat{S}(t)}\hat{x}_k(t)+\underbrace{r_k}_{\text{negligible}}$$
$$\approx\frac{m_k\hat{S}(t)}{a_k+\hat{S}(t)}\hat{x}_k(t),$$
$$\beta_{k,-1}\Big(\hat{S}(t),\hat{x}(t)\Big)=d_k(t)\hat{x}_k(t)=(D+\mu_k)\hat{x}_k(t),$$

and $b_k(t)$ and $d_k(t)$ are the birth and death rates of kth species at time t.

Over compact time intervals, the immigration rate of the species density is a negligible term, thus this stochastic process behaves the same as in the model in the absence of immigration given in Chapter 3. The diffusion approximation over compact time intervals is still the Ornstein-Uhlenbeck process with expression:

$$\mathrm{d}\hat{S}(t)=\Big\{(-D-\sum\frac{\gamma_kx_k^*}{y_k})(\hat{S}-S^*)-\sum\frac{b_k^*}{y_k}(\hat{x}_k-x_k^*)\Big\}\mathrm{d}t \quad (5.6)$$
$$\mathrm{d}\hat{x}_k(t)=\gamma_kx_k^*(\hat{S}-S^*)\mathrm{d}t+\sqrt{b_k^*x_k^*}\mathrm{d}W_{k,1}-\sqrt{d_kx_k^*}\mathrm{d}W_{k,-1}. \quad (5.7)$$

where $\gamma_k = \frac{\mathrm{d}b_k}{\mathrm{d}S}|_{S^*}$ and the equilibrium state (S^*, x_1^*, x_2^*) is only dependent on the initial states.

5.3.2 Stochastic Model at a Long Time Scale

As discussed in Chapter 3, the law of large numbers and the central limit theorem are applicable over compact time intervals, but will not hold over an infinite time interval. Actually, over an infinite interval, the absorption state will no longer be reached given even a small immigration rate. Therefore it is necessary to derive a new diffusion approximation at the long time scale, where the immigration rate $R_k = Vr_k$ will reveal its role as a significant mechanism in the dynamics.

Similarly to Sections 3.5 and 3.6, we define the population density for large V at longer timescales as,

$$\begin{aligned}
\hat{s}(t) &\stackrel{\text{def}}{=} \hat{S}(Vt), \\
\hat{w}(t) &\stackrel{\text{def}}{=} \frac{1}{V}\hat{N}(Vt), \\
\hat{w}(0) &= \bar{w}(0) = w, \\
\hat{S}(0) &= s.
\end{aligned}$$

The stochastic process at the long time scale is then defined by the expres-

sion:

$$\mathrm{d}\hat{S} = V\Big\{\Big(S^{\mathrm{in}} - \hat{S}(t)\Big)D - \sum_{k=2}^{n} \frac{\hat{w}_k(t)}{y_k}\frac{m_k\hat{S}(t)}{a_k + \hat{S}(t)}\Big\}\mathrm{d}t,$$

$$\mathrm{d}\hat{w}_k = \Big\{R_k + V\Big(\frac{m_k\hat{S}(t)}{a_k + \hat{S}(t)} - (D + \mu_k)\Big)\hat{w}_k(t)\Big\}\mathrm{d}t + \sum_{l} l\beta_{k,l}(\hat{w})^{\frac{1}{2}}\mathrm{d}B_{k,l}.$$

5.3.2.1 Projection Map onto the Deterministic Trajectory

Projecting the stochastic process onto the deterministic trajectory as $t \to \infty$, we have the projection map with definition,

$$\begin{aligned}\pi(\hat{w}) &= \Big(\pi_0(\hat{w}), \pi_1(\hat{w}), \pi_2(\hat{w})\Big) \\ &= \Big(\lim_{t\to\infty}\bar{S}(t,\hat{w}), \lim_{t\to\infty}\bar{x}_1(t,\hat{w}), \lim_{t\to\infty}\bar{x}_2(t,\hat{w})\Big).\end{aligned}$$

As was discussed in Subection 3.5.2, the projection map can well approximate the stochastic process at the long time scale, namely $\hat{w} \Rightarrow \pi(\hat{w})$.

Since $R_k = o(V)$, there is no change in the derivatives of the projection

map. That is,

$$
\begin{aligned}
\frac{\partial \pi_1}{\partial w_1} &= \frac{\gamma_2 \kappa_2 \pi_2}{\gamma_1 \kappa_1 \pi_1 + \gamma_2 \kappa_2 \pi_2} \\
\frac{\partial \pi_2}{\partial w_1} &= -\frac{\kappa_1}{\kappa_2} \frac{\partial \pi_1}{\partial w_1} \\
\frac{\partial \pi_1}{\partial w_2} &= -\frac{\gamma_1 \kappa_2 \pi_1}{\gamma_1 \kappa_1 \pi_1 + \gamma_2 \kappa_2 \pi_2} \\
\frac{\partial \pi_2}{\partial w_2} &= \frac{\partial \pi_1}{\partial w_2} \\
\frac{\partial^2 \pi_1}{\partial w_1^2} &= \frac{-\kappa_1 \kappa_2 \gamma_1 \gamma_2 \pi_2}{(\kappa_1 \pi_1 \gamma_1 + \kappa_2 \pi_2 \gamma_2)^2} \Big(1 + \frac{\gamma_2(\kappa_1 \pi_1 + \kappa_2 \pi_2)}{\kappa_1 \pi_1 \gamma_1 + \kappa_2 \pi_2 \gamma_2}\Big), \\
\frac{\partial^2 \pi_2}{\partial w_1^2} &= -\frac{\kappa_1}{\kappa_2} \frac{\partial^2 \pi_1}{\partial w_1^2}, \\
\frac{\partial^2 \pi_1}{\partial w_2^2} &= \frac{\kappa_2^2 \gamma_1 \gamma_2 \pi_1}{(\kappa_1 \pi_1 \gamma_1 + \kappa_2 \pi_2 \gamma_2)^2} \Big(1 + \frac{\gamma_1(\kappa_1 \pi_1 + \kappa_2 \pi_2)}{\kappa_1 \pi_1 \gamma_1 + \kappa_2 \pi_2 \gamma_2}\Big), \\
\frac{\partial^2 \pi_2}{\partial w_2^2} &= -\frac{\kappa_1}{\kappa_2} \frac{\partial^2 \pi_1}{\partial w_2^2}.
\end{aligned}
$$

It turns out that due to the significant influence of the immigration rate, which will be explored later in this chapter, the impact of the inaccuracy of the derivatives of the projection map (absence of the higher order terms of the Taylor expansion of the deterministic trajectory) is greatly weakened. Therefore, the derivatives of the projection map derived above are sufficient to calculate the final stationary distribution. It is not necessary to correct the derivatives of the projection map as we did in Subsection 3.5.3.1 by considering the higher order terms in the Taylor expansion of the deterministic trajectory.

5.3.2.2 Diffusion Approximation at the Long Time Scale

As before we can apply Itô's transformation to derive the approximation of the projection map, which is

$$\pi_k\Big(\hat{w}(t)\Big) = \pi_k\Big(\hat{w}(0)\Big) + \underbrace{\int_0^t (D\pi_k)\Big(\hat{w}(s)\Big)\cdot F\Big(\hat{w}(s)\Big)ds}_{\text{Finite Variation Component}} + \underbrace{\sum_i^2 \int_0^t (\partial_i \pi_k)\Big(\hat{w}(s)\Big)dM_i(s)}_{\text{Martingale Component}}$$
$$+ \underbrace{\frac{1}{2}\sum_i^2 \int_0^t (\partial_i^2 \pi_k)\Big(\hat{w}(s)\Big)d[M_i]_s}_{\text{Quadratic Variation Component}} + \epsilon(t). \tag{5.8}$$

The martingale component and the quadratic variation component are the same as in the model in the absence of immigration, but the finite variation component is not. This is because the immigration term R_k in the $F(\hat{w}(t))$ contributes an extra non-zero value compared with the model in the absence of immigration,

$$\begin{aligned}(D\pi_k)\Big(\hat{w}(s)\Big)\cdot F\Big(\hat{w}(s)\Big) &= (D\pi_k)\Big(\hat{w}(s)\Big)\cdot\Big(R_k + V\gamma_k(\hat{S}-S^*)\hat{w}_k\Big)\\ &= (D\pi_k)\Big(\hat{w}(s)\Big)\cdot R_k + 0\\ &= R_1\frac{\mathrm{d}\pi_k}{\mathrm{d}\hat{w}_1} + R_2\frac{\mathrm{d}\pi_k}{\mathrm{d}\hat{w}_2}\end{aligned}$$

With the help of the weak convergence[37], $\pi_k\Big(\hat{w}(t)\Big)$ converges weakly to $\Pi(w,t) = \Big(\Pi_1(w,t), \Pi_2(w,t)\Big)$ with expression,

$$
\begin{aligned}
d\Pi_1 &= \Big\{R_1 \cdot \mathrm{d}_1\Pi_1 + R_2 \cdot \mathrm{d}_2\Pi_1 + (\mathrm{d}_1^2\Pi_1)d_1\Pi_1 + (\mathrm{d}_2^2\Pi_1)d_2\Pi_2\Big\}dt \\
&\quad + (\mathrm{d}_1\Pi_1)\sqrt{d_1\Pi_1}\Big(dW_{1,1} - dW_{1,-1}\Big) + (\mathrm{d}_2\Pi_1)\sqrt{d_2\Pi_2}\Big(dW_{2,1} - dW_{2,-1}\Big) \\
d\Pi_2 &= \Big\{R_1 \cdot \mathrm{d}_1\Pi_2 + R_2 \cdot \mathrm{d}_2\Pi_2 + (\mathrm{d}_1^2\Pi_2)d_1\Pi_1 + (\mathrm{d}_2^2\Pi_2)d_2\Pi_2\Big\}dt \\
&\quad + (\mathrm{d}_1\Pi_2)\sqrt{d_1\Pi_1}\Big(dW_{1,1} - dW_{1,-1}\Big) + (\mathrm{d}_2\Pi_2)\sqrt{d_2\Pi_2}\Big(dW_{2,1} - dW_{2,-1}\Big).
\end{aligned}
$$

This is the diffusion approximation for the stochastic process of the species densities with immigration present at the long time scale.

5.3.3 Stationary Distribution of the Relative Abundance of the First Species

In the model with two quasi-neutral species, $p(t) = \frac{\Pi_1(t)}{\Pi_1(t)+\Pi_2(t)}$ is a diffusion on [0,1], and the dynamics of the relative abundance of the first species is sufficient to explain the whole population dynamics. Since $\Big(\Pi_1(t), \Pi_2(t)\Big)$ is kept on the centre manifold, the total density approximates to,

$$\Theta = \Pi_1 + \Pi_2 = \frac{(S^{\mathrm{in}} - S^*)D}{\kappa_1 p + \kappa_2(1-p)}.$$

As discussed in Subsection 3.6.2, by applying Itô's formula, the diffusion of the relative abundance of the first species $p(t)$ is expressed as,

$$\begin{aligned} dp &= \frac{\mathrm{d}p}{\mathrm{d}\Pi_1}(\mathrm{d}\Pi_1) + \frac{1}{2}\frac{\mathrm{d}^2 p}{\mathrm{d}\Pi_1^2}(\mathrm{d}\Pi_1)^2 \\ &= \frac{\kappa_2(1-p)+\kappa_1 p}{\kappa_2\Theta}(\mathrm{d}\Pi_1) + \frac{(\kappa_2(1-p)+\kappa_1 p)(\kappa_1-\kappa_2)}{\kappa_2^2\Theta^2}(\mathrm{d}\Pi_1)^2 \\ &= b(p)dt + a(p)dB_t, \end{aligned}$$

with drift coefficient

$$\begin{aligned} b(p) = \frac{\kappa_2(1-p)+\kappa_1 p}{\kappa_2}\Big\{ &\frac{R_1}{\Theta}\mathrm{d}_1\Pi_1 + \frac{R_2}{\Theta}\mathrm{d}_2\Pi_1 + d_1 p\mathrm{d}_1^2\Pi_1 + d_2(1-p)\mathrm{d}_2^2\Pi_1 \\ &+ \frac{2(\kappa_1-\kappa_2)}{\kappa_2\Theta}\Big(d_1 p(\mathrm{d}_1\Pi_1)^2 + d_2(1-p)(\mathrm{d}_2\Pi_1)^2\Big)\Big\} \end{aligned}$$

and diffusion coefficient $a(p)$ satisfying

$$\frac{1}{2}a(p)^2 = \frac{\Big(\kappa_2(1-p)+\kappa_1 p\Big)^2}{\kappa_2^2\Theta}\Big\{d_1 p(\mathrm{d}_1\Pi_1)^2 + d_2(1-p)(\mathrm{d}_2\Pi_1)^2\Big\}.$$

Using the forward Kolmogorov equation, we have that the probability density of the relative abundance of the first species $f(t,p)$ satisfies,

$$\frac{\mathrm{d}f(t,p)}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}p}\{b(p)f(t,p)\} + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}p^2}\{a^2(p)f(t,p)\}. \tag{5.9}$$

To derive the stationary distribution, we integrate Eq.(5.9) throughout formally with respect to p, giving:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(1-F(t,p)\Big) = b(p)f(t,p) - \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}p}\Big(a^2(p)f(p)\Big), \tag{5.10}$$

with $F(t,p)$ as the distribution function:

$$F(t,p) = \int_0^p f(t,y)dy.$$

Since the stationary distribution is independent of t, we define it as $f(p)$, and the left hand side of Eq.(5.10) equals zero,

$$0 = b(p)f(p) - \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}p}\Big(a^2(p)f(p)\Big).$$

Then the stationary probability density is derived in the form of

$$f(p) = \frac{\text{const}}{a^2(p)} \exp\Big\{2\int_0^p \frac{b(y)}{a^2(y)}dy\Big\},$$

with the constant satisfying $\int_0^1 f(p) = 1$.

To expand the expression in detail, we substitute the derivatives of the

projection map into $b(p)$ and $a^2(p)$, obtaining

$$
\begin{aligned}
2\frac{b(p)}{a(p)^2} &= \frac{2(\kappa_1-\kappa_2)}{(\kappa_1-\kappa_2)p+\kappa_2} \\
&+ \frac{\Big(\gamma_1\kappa_1 p+\gamma_2\kappa_2(1-p)\Big)\Big(R_1\gamma_2(1-p)-R_2\gamma_1 p\Big)}{\Big((\kappa_1-\kappa_2)p+\kappa_2\Big)p(1-p)\Big((d_2\gamma_1^2-d_1\gamma_2^2)p+d_1\gamma_2^2\Big)} \\
&+ \frac{\gamma_1\gamma_2(d_2\kappa_2-d_1\kappa_1)}{\Big((\kappa_1-\kappa_2)p+\kappa_2\Big)\Big((d_2\gamma_1^2-d_1\gamma_2^2)p+d_1\gamma_2^2\Big)} \\
&+ \frac{\gamma_1\gamma_2(d_2\kappa_2\gamma_1-d_1\kappa_1\gamma_2)}{\Big((\kappa_1\gamma_1-\kappa_2\gamma_2)p+\kappa_2\Big)\Big((d_2\gamma_1^2-d_1\gamma_2^2)p+d_1\gamma_2^2\Big)} \\
&= \Big\{2+\frac{(\gamma_1\kappa_1-\gamma_2\kappa_2)(\gamma_1R_2+\gamma_2R_1)-\gamma_1\gamma_2(d_1\kappa_1-d_2\kappa_2)}{\kappa_1d_1\gamma_2^2-\kappa_2d_2\gamma_1^2} \\
&- \frac{\kappa_1-\kappa_2}{\kappa_1d_1\gamma_2^2-\kappa_2d_2\gamma_1^2}(R_2\gamma_1^2+R_1\gamma_2^2)\Big\}\frac{(\kappa_1-\kappa_2)}{(\kappa_1-\kappa_2)p+\kappa_2} \\
&+ \Big\{1-\frac{(\gamma_1\kappa_1-\gamma_2\kappa_2)(\gamma_1R_2+\gamma_2R_1)-\gamma_1\gamma_2(d_1\kappa_1-d_2\kappa_2)}{\kappa_1d_1\gamma_2^2-\kappa_2d_2\gamma_1^2} \\
&+ \frac{d_2\gamma_1^2-d_1\gamma_2^2}{\kappa_1d_1\gamma_2^2-\kappa_2d_2\gamma_1^2}\Big(\frac{R_2\kappa_1}{d_2}+\frac{R_1\kappa_2}{d_1}\Big)\Big\}\frac{d_2\gamma_1^2-d_1\gamma_2^2}{(d_2\gamma_1^2-d_1\gamma_2^2)p+d_2\gamma_1^2} \\
&- \frac{\kappa_1\gamma_1-\kappa_2\gamma_2}{(\kappa_1\gamma_1-\kappa_2\gamma_2)p+\kappa_2\gamma_2}-\frac{R_2}{d_2}\frac{1}{1-p}+\frac{R_1}{d_1}\frac{1}{p} \\
a(p)^2 &= \frac{p(1-p)\Big((d_2\gamma_1^2-d_1\gamma_2^2)p+d_2\gamma_1^2\Big)\Big((\kappa_1-\kappa_2)p+\kappa_2\Big)^3}{(S^{\text{in}}-S^*)D\Big((\kappa_1\gamma_1-\kappa_2\gamma_2)p+\kappa_2\gamma_2\Big)^2}.
\end{aligned}
$$

with $\kappa=\frac{d_k}{y_k}$, $\gamma_k=\frac{\mathrm{d}b_k}{\mathrm{d}S}|_{S^*}=\frac{(m_k-d_k)^2}{m_ka_k}$.

Therefore, the final stationary distribution of the relative abundance of

the first species is derived in the following expression

$$\begin{aligned} f(p) = \text{const} \cdot & \Big((\kappa_1-\kappa_2)p+\kappa_2\Big)^{\frac{(\gamma_1-\gamma_2)(\gamma_1\kappa_2R_2+\gamma_2\kappa_1R_1)-\gamma_1\gamma_2(d_1\kappa_1-d_2\kappa_2)}{\kappa_1 d_1\gamma_2^2-\kappa_2 d_2\gamma_1^2}-1} \\ & \cdot \Big((\kappa_1\gamma_1-\kappa_2\gamma_2)p+\kappa_2\gamma_2\Big) \\ & \cdot \Big((d_2\gamma_1^2-d_1\gamma_2^2)p+d_1\gamma_2^2\Big)^{\frac{(\frac{R_2}{d_2}\gamma_2+\frac{R_1}{d_1}\gamma_1)(\kappa_2\gamma_1 d_2-\kappa_1\gamma_2 d_1)-\gamma_1\gamma_2(d_1\kappa_1-d_2\kappa_2)}{\kappa_1 d_1\gamma_2^2-\kappa_2 d_2\gamma_1^2}} \\ & \cdot p^{\frac{R_1}{d_1}-1}(1-p)^{\frac{R_2}{d_2}-1} \end{aligned} \tag{5.11}$$

with the constant value satisfying $\int_0^1 f(p) = 1$.

As was assumed in the previous discussion, the intrinsic death rate is considered to be negligible, that is $d_1 = d_2 = D$. Then the expression for the stationary distribution of the relative abundance of the first species becomes:

$$\begin{aligned} f(p) = \text{const} \cdot & \underbrace{\Big((\kappa_1-\kappa_2)p+\kappa_2\Big)^{-\frac{R_1}{D}+\frac{(R_1-R_2)(\gamma_1-\gamma_2)\gamma_1\kappa_2}{D(\gamma_1^2\kappa_2-\gamma_2^2\kappa_1)}-\frac{\gamma_1\gamma_2(\kappa_1-\kappa_2)}{\gamma_1^2\kappa_2-\gamma_2^2\kappa_1}(\frac{R_1}{D}-1)-1}}_{①} \\ & \underbrace{\Big((D\gamma_1^2-D\gamma_2^2)p+D\gamma_2^2\Big)^{-\frac{R_2}{D}-\frac{(R_1-R_2)(\gamma_1-\gamma_2)\gamma_1\kappa_2}{D(\gamma_1^2\kappa_2-\gamma_2^2\kappa_1)}+\frac{\gamma_1\gamma_2(\kappa_1-\kappa_2)}{\gamma_1^2\kappa_2-\gamma_2^2\kappa_1}(\frac{R_1}{D}+1)}}_{②} \\ & \underbrace{\Big((\kappa_1\gamma_1-\kappa_2\gamma_2)p+\kappa_2\gamma_2\Big)}_{③} \cdot p^{\frac{R_1}{D}-1}\cdot(1-p)^{\frac{R_2}{D}-1}. \end{aligned} \tag{5.12}$$

Neutral Model

In the strictly neutral case where $\gamma_1 = \gamma_2,\ \kappa_1 = \kappa_2,\ d_1 = d_2 = D$, the relative abundance of the first species p is beta distributed with probability

density,

$$f(p) = \frac{\Gamma(\frac{R_1+R_2}{D})}{\Gamma(\frac{R_1}{D})\Gamma(\frac{R_2}{D})} p^{\frac{R_1}{D}-1}(1-p)^{\frac{R_2}{D}-1}. \tag{5.13}$$

The mean value of this beta distribution equals the relative abundance of the first species in the meta-community $\frac{R_1}{R_1+R_2} = q_1$. The value of the variance is $\frac{R_1 R_2 D}{(R_1+R_2)^2(R_1+R_2+D)}$.

5.3.4 Effects of the Parameters on the Stationary Distribution

Compared with the stationary distribution of the strictly neutral model given in Eq. (5.13), which is only dependent on the immigration rate and dilution rate, the stationary distribution of the quasi-neutral model in Eq. (5.12) has a selection term (①, ②, ③ in Eq. (5.12)) which moves the stationary distribution of the relative abundance away from the beta distribution. This selection term is dependent on the immigration rate R_k, the sensitivity of the birth rate to the resource at the equilibrium state γ_k (trade-offs parameter), the yields y_k and the death rate d_k, and is too complicated to expand by Taylor expansion in $\Delta(\kappa)$ and $\Delta(\gamma)$. Therefore to analyse the impact of the parameters, we plot the analytical stationary distributions in Figure 5.1 over different parameters.

5.3.4.1 Effects of the Immigration Rate

In each panel of Figure 5.1, the distributions of the relative abundances show large sensitivities to small changes in the immigration rate. It is apparent that the species with a higher abundance in the meta-community has the

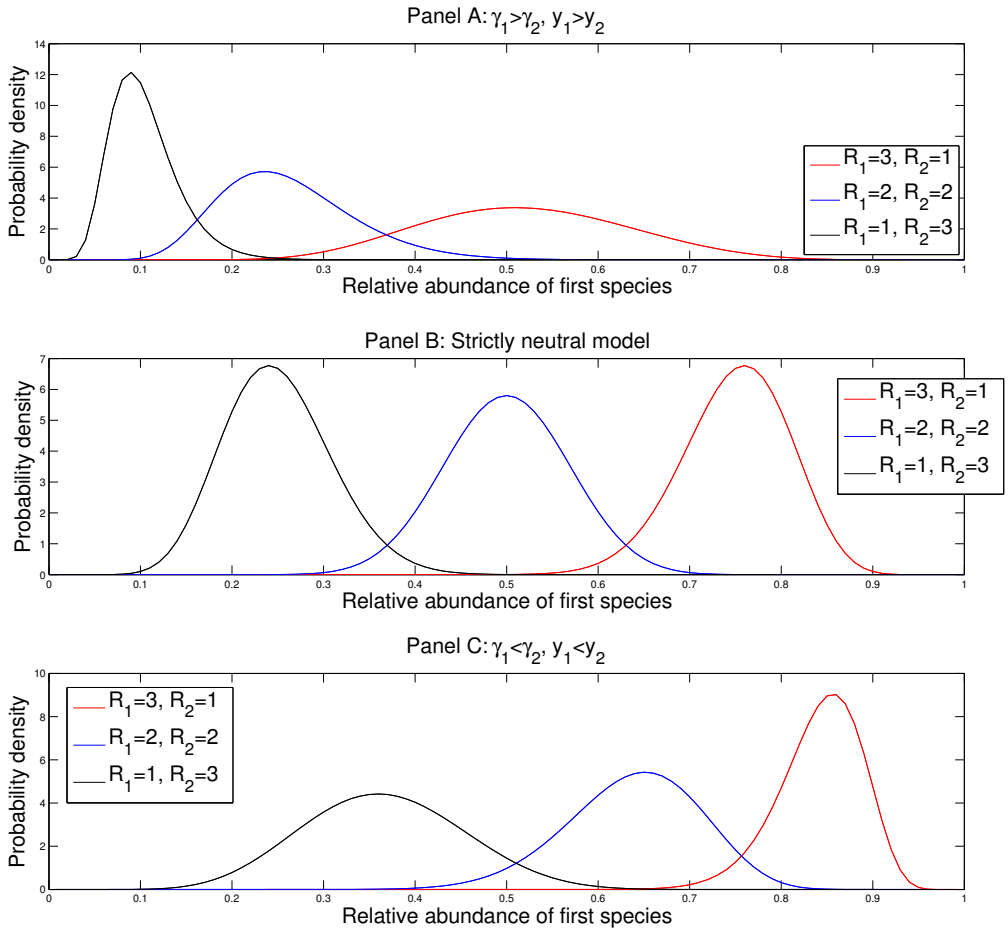

Figure 5.1: The dependence of the stationary distributions of the relative frequency of the first species on different immigration rates R. The ratios of $\frac{\gamma_1}{\gamma_2}$ are 3, 1 and $\frac{1}{2}$, and ratios of $\frac{\kappa_1}{\kappa_2}$ are $\frac{1}{10}$, 1 and 3 in Panel A, B and C respectively.

advantage of obtaining a larger relative abundance in the local community. Immigration, even at small rates, plays a significant role in determining the community's diversity.

5.3.4.2 Effects of y_k and γ_k

To remove the impact of the immigration rate, here we assume that $R_1 = R_2 = 2$. The blue lines in Figure 5.1 shows stationary distributions of the relative abundance of the first species with different values of the ratios $\frac{\gamma_1}{\gamma_2}$ and $\frac{\kappa_1}{\kappa_2}$, where $\kappa_k = \frac{D}{y_k}$.

Given the parameter relationships $\frac{\gamma_1}{\gamma_2} = 3, \frac{\kappa_1}{\kappa_2} = \frac{1}{10}$, the term ① in Eq.(5.12) shows a selective advantage for the first species (which skews the distribution to the right), but terms ② and ③ play opposite roles. The result is that the mean of the blue line in Figure 5.1 A with $\frac{\gamma_1}{\gamma_2} = 3, \frac{y_1}{y_2} = 10$ is less than the mean of the strictly neutral model in Figure 5.1 B, which shows that larger γ_k and y_k reduce the relative abundance of the first species.

From this example, we can see that it is hard to determine the impact of a single parameter on the stationary distribution of the relative abundance. Instead, what may be analysed is the impact of a single parameter given the condition that all other parameters are identical for both quasi-neutral species.

Effects of the Yields y_k with Identical γ_k

If we assume γ_1 is identical to γ_2, then the expression for the stationary distribution of the relative abundance of the first species may be derived as

$$\begin{aligned} f(p) &= \text{const} \cdot \Big(1 + (\frac{\kappa_1}{\kappa_2} - 1)p\Big)^{-1} \cdot p^{\frac{R_1}{D}-1} \cdot \Big(1-p\Big)^{\frac{R_2}{D}-1} \\ &= \text{const} \cdot \Big(1 + (\frac{y_2}{y_1} - 1)p\Big)^{-1} \cdot p^{\frac{R_1}{D}-1} \cdot \Big(1-p\Big)^{\frac{R_2}{D}-1}. \end{aligned} \tag{5.14}$$

In Eq. (5.14), the selection term $\left(1+(\frac{y_2}{y_1}-1)p\right)^{-1}$ gives the species with larger y_k an advantage in obtaining a higher relative abundance. This result can be demonstrated in Figure 5.2 where the stationary distributions of the relative abundance of the first species with different values of the ratio $\frac{y_1}{y_2}$ are plotted. However, the discrepancies among the stationary distributions with different yields ratios $\frac{y_1}{y_2}$ are not apparent in both top and bottom figures in Figure 5.2 where the relative abundance in the immigrating populations are different.

In the local microbial community, even with low immigration rate, the exact number R_k will be much larger. In Figure 5.3, after increasing the immigration rate from 4 to 40 (this value is still negligible compared to the large local microbial population size), the discrepancies among the distributions over different ratios $\frac{y_1}{y_2}$ is hard to pinpoint. This may be explained by the fact that, in Eq. (5.14), the power of the selective term $\left(1+(\frac{y_2}{y_1}-1)p\right)$ is -1, which is far less than the power $\frac{R_1}{D}$ of p and power $\frac{R_2}{D}$ of $1-p$, and this results in a negligible influence on the final analytical prediction.

Therefore, we can confidently conclude that the yields y_k do not affect the stationary distribution of the relative abundance greatly if the trade-offs parameters γ_k are identical for both species.

Effects of γ_k with Identical Yields y_k

Similarly, in this subsection, we will test the impact of the trade-off parameter, γ_k, on the stationary distribution of the relative abundance of the first species.

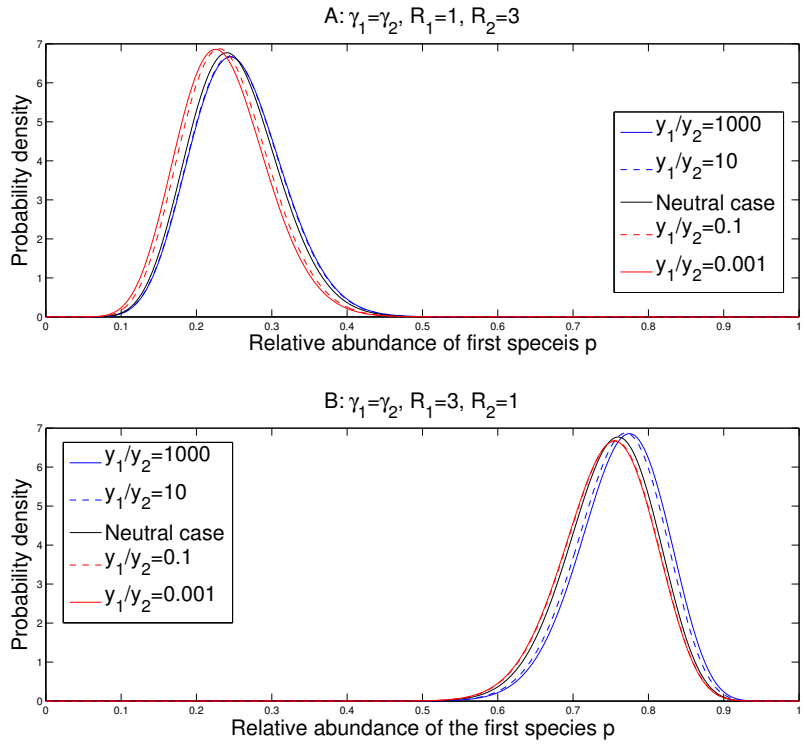

Figure 5.2: The analytical stationary distributions of the relative frequency of the first species with different yields y_k but identical γ_k. The immigration population per unit time in the top figure are $R_1 = 1$, $R_2 = 3$, and in the bottom figure are $R_1 = 3$, $R_2 = 1$.

By assuming $y_1 = y_2$, the expression of the stationary distribution of the relative abundance of the first species is,

$$f(p) = \text{const} \cdot \Big((\gamma_1 - \gamma_2)p + \gamma_2\Big)\Big((\gamma_1^2 - \gamma_2^2)p + \gamma_2^2\Big)^{-\frac{R_2\gamma_2 + R_1\gamma_1}{D(\gamma_1+\gamma_2)}} p^{\frac{R_1}{D}-1}(1-p)^{\frac{R_2}{D}-1} \tag{5.15}$$

As was discussed in the last subsection, the first term in Eq. (5.15)

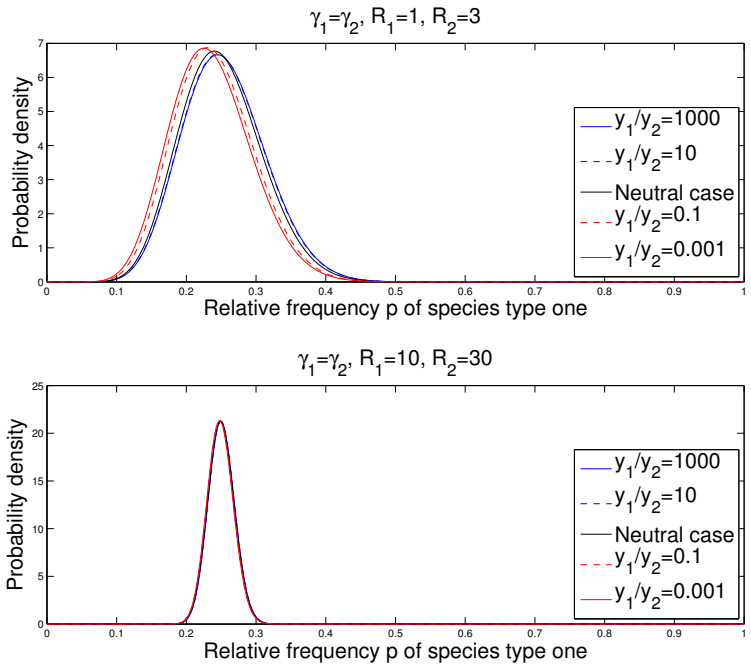

Figure 5.3: The effects of the yields on the stationary distributions of the relative abundance of the first species for different immigration rates. In both figures, $\frac{\gamma_1}{\gamma_2} = 1$ and $\frac{R_1}{R_2} = \frac{1}{3}$, but the immigration population size per unit time in the top and bottom figures are 4 and 40 respectively.

plays a small role in the final result due to the small value of its power, 1. However, the second term cannot be ignored due to the larger value of its power, $-\frac{R_2\gamma_2+R_1\gamma_1}{D(\gamma_1+\gamma_2)}$. Since this power is always negative, the second term then represents a selective disadvantage to the species with larger γ. These effects are illustrated in Figure 5.4, where larger γ_k decreases the mean of the relative abundance of this species, and *vice versa*.

To determine whether increasing the immigration rates will change the effect of γ_k, we plot the variation in the stationary distributions for different

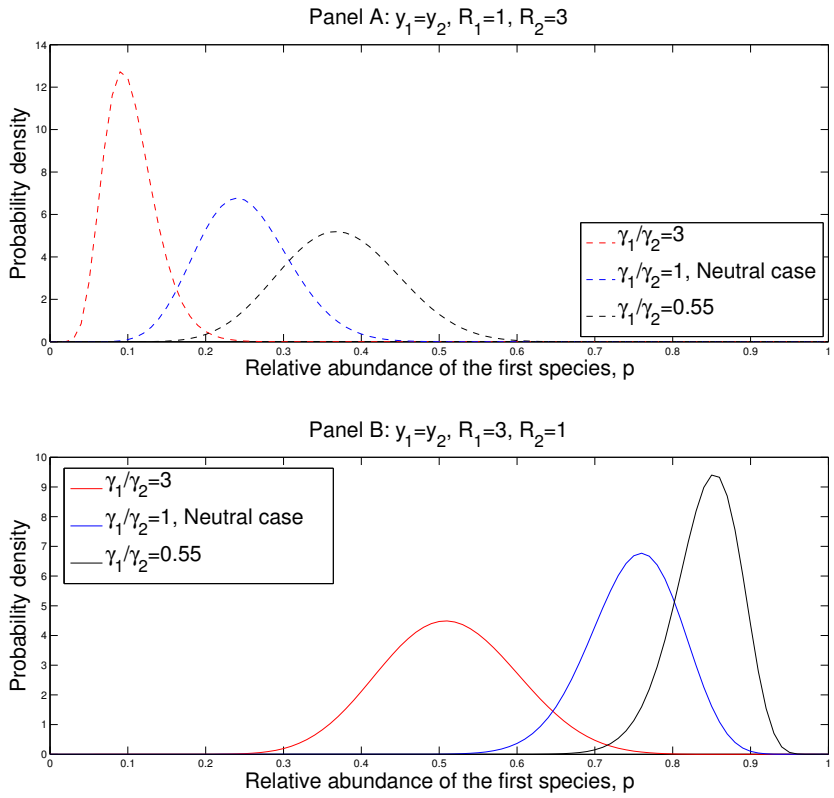

Figure 5.4: The stationary distributions of the relative abundance of the first species with different γ_k but same yields. The lines and dots in red are results of the model with parameters: $a_1 = 1 \cdot 10^{-4}$, $m_1 = 0.15$, $a_2 = 0.2 \cdot 10^{-4}$, $m_2 = 0.09$, $y_1 = y_2 = 5 \cdot 10^7$. Blue lines and dots represent the results of neutral model with:$a_1 = a_2 = 1 \cdot 10^{-4}$, $m_1 = 0.15$, $y_1 = y_2 = 5 \cdot 10^7$. The black lines and dots gives results of the models with parameter values: $a_1 = 1 \cdot 10^{-4}$, $m_1 = 0.15$, $a_2 = 1 \cdot 10^{-3}$, $m_2 = 0.825$, $y_1 = y_2 = 5 \cdot 10^7$. The models In Panel A have immigration rates: $R_1 = 1$, $R_2 = 3$. The immigration rates of the models in Panel B are: $R_1 = 3$, $R_2 = 1$.

γ_k under different immigration rates in Figure 5.5. This illustrates that the mean of each distribution does not obviously change with the increasing R, but the standard deviation greatly decreases which indicates a weakened

role of stochastic drift.

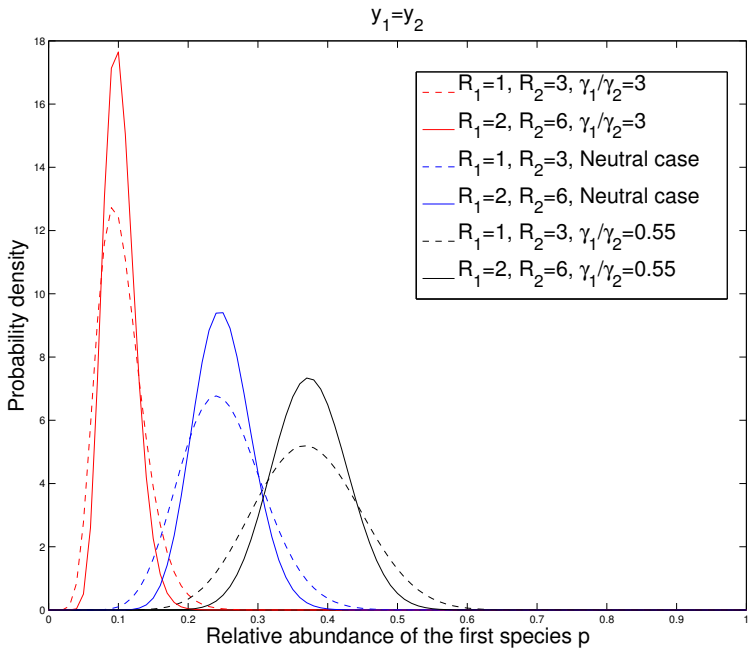

Figure 5.5: The effects of γ on the stationary distributions of the relative frequency of the first species with different size of immigration population. The lines and dots in red are the results of the models with parameters: $a_1 = 1 \cdot 10^{-4},\ m_1 = 0.15,\ a_2 = 0.2 \cdot 10^{-4},\ m_2 = 0.09,\ y_1 = y_2 = 5 \cdot 10^7$. Blue lines and dots represent the results of the strict neutral models with: $a_1 = a_2 = 1 \cdot 10^{-4},\ m_1 = m_2 = 0.15,\ y_1 = y_2 = 5 \cdot 10^7$. The lines and dots in black give the results of the models with parameter values: $a_1 = 1 \cdot 10^{-4},\ m_1 = 0.15,\ a_2 = 1 \cdot 10^{-3},\ m_2 = 0.825,\ y_1 = y_2 = 5 \cdot 10^7$. The models in line has immigration rates $R_1 = 2$ ans $R_2 = 6$, which is twice of the immigration rates in the models in dots.

Therefore, ignoring y_k, the species with smaller γ_k has an advantage in reaching a higher relative abundance in the stationary phase. This result demonstrate that the quasi-neutral species with different trade-off parameters exhibit substantial differences.

In the following section, the analytical stationary distributions will be compared with the results of numerical calculations.

5.4 Numerical Models

The simulation of the model in the presence of immigration is almost the same as explained in Figure 4.1. The only difference is that a small immigration rate will be added into the birth rate for each species, which ensures the existence of birth events even when the current population is zero for both species. After running for a long enough time (400000 time units), the probability density of the relative abundance of the first species may be estimated using the Kernel smoothing function estimating method (command ksdensity in Matlab). In Figure 5.6, the results of the numerical calculations with immigration rate $R_1 = R_2 = 2$ are plotted together with the analytical approximations.

In Figure 5.6, all the analytical approximated stationary distributions fit well with the numerically estimated stationary distributions. With identical yields, the result from the last section is proved that the mean of the stationary distribution increases with the decrease of γ_k, which indicates that the species with smaller trade-off parameter has an advantage in reaching a higher relative abundance over long time scales.

5.5 Comparison With Hubbell's Neutral Model

Hubbell's neutral model is built on a limited set of assumptions and unreasonably assumes neutrality at the individual level. In most cases, it

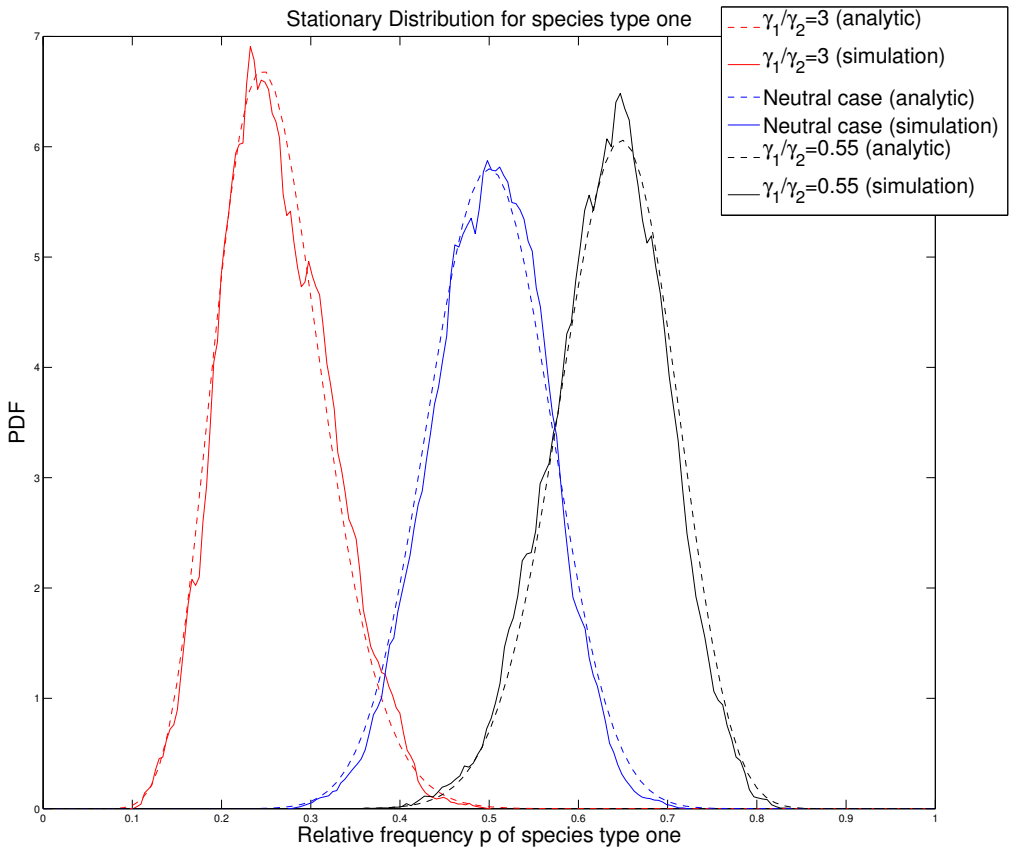

Figure 5.6: The stationary distributions of the relative abundance of the first species with different γ values but the same yields, and $R_1 = R_2 = 2$. The lines are the results of numerical calculations and dots are the analytical predictions. The dots and lines in red represent the models with parameters: $a_1 = 1{\cdot}10^{-4}$, $a_2 = 2{\cdot}10^{-5}$, $m_1 = 0.15$, $m_2 = 0.09$, $y_1 = y_2 = 5{\cdot}10^7$. Blue lines and dots are the strict neutral models with parameters: $a_1 = a_2 = 1 \cdot 10^{-4}$, $m_1 = m_2 = 0.15$, $y_1 = y_2 = 5 \cdot 10^7$. The dots and lines in black give the results of the models with parameters: $a_1 = 1 \cdot 10^{-4}$, $a_2 = 1 \cdot 10^{-3}$, $m_1 = 0.15$, $m_2 = 0.825$, $y_1 = y_2 = 5 \cdot 10^7$.

has been rejected by statistical tests [60][42][43]. However, the useful point is that it can be considered as a null model for community assembly to compare with actual data and assess the influence of parameters [66].

In this section, Hubbell's Unified Neutral Theory of Biodiversity will be calibrated for the chemostat community in Subsection 5.5.1, to compare with the our niche structured quasi-neutral demographic model in Subsection 5.5.2.

5.5.1 Hubbell's Neutral Model

Hubbell's Unified Neutral Theory of Biodiversity has been introduced in Subsection 1.3.1.1. In this theory, all individuals are assumed to be ecologically identical. The research focus in this thesis is on the local community dynamics, and assumes that there is a meta-community coupled to the island chemostat community with known relative abundances q_i and population size Θ_m.

In the model, the local community is saturated with a total of Θ individuals. If there is a change in the community, an individual must die, and the dead individual is immediately replaced by an individual immigrating from the meta-community, with probability m, or by a newborn individual which is the offspring of a random individual within the local community with probability $1 - m$ [31].

To explain the local chemostat environment by neutral theory, instead of directly testing Hubbell's neutral model, we focus on Sloan's near-neutral model described in Ref [62]. This near-neutral model is built on the core concepts of Hubbell's neutral theory, but makes two main advances. Firstly, a

competitive advantage or disadvantage is allowed for each species. Secondly, the near-neutral model translates Hubbell's discrete model into a continuous diffusion equation so that the large population-sized microbial environment may be modelled.

Then based on the expression for Hubbell's neutral model in Eqs. (1.2) and (1.3) in Chapter 1, the effective birth and death rates for each species are encapsulated in the following expressions ,

$$Pr(N_i + 1/N_i) = (\frac{\Theta - N_i}{\Theta})\Big(mq_i + (1+\alpha_i)(1-m)(\frac{N_i}{\Theta - 1})\Big)$$
$$Pr(N_i - 1/N_i) = (\frac{N_i}{\Theta})\Big(m(1-q_i) + (1-\alpha_i)(1-m)(\frac{\Theta - N_i}{\Theta - 1})\Big),$$

where q_i is the relative abundance in the meta-community, m is the migration rate, Θ is the total population in the local community and α_i is the advantage parameter for each species [62].

Assuming there are two species in the model, let $p = \frac{N_1}{\Theta}$ be the relative abundance of the first species in the local community. Sloan shows the probability density function of p, $f(p,t)$, is given by the forward Kolmogorov equation [62]:

$$\frac{\mathrm{d}f(p,t)}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}p}\Big(Mf\Big) + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}p}\Big(Vf\Big),$$

where, M, is the expected rate of change in frequency, approximates to

$$\frac{m(q_1 - p) + (1-m)2\alpha_i p(1-p)}{\Theta},$$

and, V, is a measure of the rate of change in variability, approximates to

$$\frac{2(1-p)p + m(q_1-p)(1-2p)}{\Theta^2}.$$

As was defined in the beginning of this chapter, the immigration rate in the local community is too small to change the deterministic dynamics, then the second term in the variance of the above process, $\frac{m(q_1-p)(1-2p)}{\Theta^2}$, is negligible. Therefore, the forward Kolmogorov equation of the probability density can be approximated as,

$$\begin{aligned}\frac{\mathrm{d}f(p,t)}{\mathrm{d}t} = &-\frac{\mathrm{d}}{\mathrm{d}p}\Big(\frac{m(q_1-p)+(1-m)2\alpha_i p(1-p)}{\Theta}f(p,t)\Big)\\ &+\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}p}\Big(\frac{2(1-p)p}{\Theta^2}f(p,t)\Big),\end{aligned}$$

where q_1 is the relative abundance of the first species in the meta-community. Then by equalising $\frac{\mathrm{d}f(p,t)}{\mathrm{d}t}$ being zero, we obtain the stationary solution for this probability density function in Sloan's near-neutral model:

$$f(p) = \text{const} \cdot p^{\Theta m q_1 - 1}(1-p)^{\Theta m(1-q_1)-1}\exp\{2\alpha_1\Theta(1-m)p\}. \tag{5.16}$$

In Hubbell's model, with strict neutrality, $\alpha_i = 0$, the density of the relative abundance of the first species in the local-community is Beta distributed with mean and variance,

$$\mu = q_1 \tag{5.17}$$

$$var = \frac{q_1(1-q_1)}{\Theta m + 1}. \tag{5.18}$$

where q_1 is the relative abundance of the first species in the meta-community.

5.5.2 Comparison

In this subsection, we will compare our quasi-neutral demographic model with the neutral model explained in Subsection 5.5.1.

We showed in Eq. (5.13) of Subsection 5.3.3 that in the strictly neutral case of our quasi-neutral demographical model, the stationary distribution of the relative abundance of the first species derived is beta distributed. Since the rate of immigration of each species is proportional to the relative abundance in the meta-community as shown in Eq. (5.1), we have the mean and variance,

$$\begin{aligned} \mu_1 &= \frac{R_1}{R_1+R_2} = q_1 \\ var_1 &= \frac{R_1 R_2 D}{(R_1+R_2)^2(R_1+R_2+D)} = \frac{q_1(1-q_1)}{\frac{R_1+R_2}{D}+1}, \end{aligned}$$

where q_1 is the relative abundance of the first species in the meta-community, and $\frac{R_1+R_2}{D}$ corresponds to the term Θm in Eq. (5.18), which is the mean of the immigration rate from the meta-community. Therefore the two beta distributions from Hubbell's neutral model and our demographic quasi-neutral model in the strictly neutral case are matched.

However, as there exists a difference between the parameters γ_k and y_k for the quasi-neutral species in the same niche, the relative abundance of our demographic quasi-neutral model depart from Hubbell's neutral model. Moreover, we cannot find a constant α_i which is independent of p that can

map our quasi-neutral model onto Sloan's model.

$$f(p) = \text{const} \cdot \underbrace{\Big((\kappa_1-\kappa_2)p+\kappa_2\Big)^{-\frac{R_1}{D}+\frac{(R_1-R_2)(\gamma_1-\gamma_2)\gamma_1\kappa_2}{D(\gamma_1^2\kappa_2-\gamma_2^2\kappa_1)}-\frac{\gamma_1\gamma_2(\kappa_1-\kappa_2)}{\gamma_1^2\kappa_2-\gamma_2^2\kappa_1}(\frac{R_1}{D}-1)-1}}_{①}$$
$$\underbrace{\Big((D\gamma_1^2-D\gamma_2^2)p+D\gamma_2^2\Big)^{-\frac{R_2}{D}-\frac{(R_1-R_2)(\gamma_1-\gamma_2)\gamma_1\kappa_2}{D(\gamma_1^2\kappa_2-\gamma_2^2\kappa_1)}+\frac{\gamma_1\gamma_2(\kappa_1-\kappa_2)}{\gamma_1^2\kappa_2-\gamma_2^2\kappa_1}(\frac{R_1}{D}+1)}}_{②}$$
$$\Big(\underbrace{(\kappa_1\gamma_1-\kappa_2\gamma_2)p+\kappa_2\gamma_2}_{③}\Big)\cdot p^{\frac{R_1}{D}-1}\cdot(1-p)^{\frac{R_2}{D}-1} \tag{5.19}$$

We have discussed the effects of parameters γ_k and y_k on the stationary distribution in the previous section and concluded that their impacts on the selective terms ①, ②, ③ in the stationary distribution Eq. (5.19) are complicated. Then a positive or negative α_i in Sloan's model cannot be determined easily by comparing Eq. (5.16) with (5.19). Nevertheless the selective terms ①, ②, ③ in Eq. (5.19) are independent of population size. This suggests that the selective parameter α_i is independent of the volume and population size of the local community. Therefore the advantage or disadvantage to the species does not change with the population size.

If the two quasi-neutral species in the model have equivalent sensitivity of the birth rate to the resource at the equilibrium state γ_k, using the results derived at the end of Subsection 5.3.4.2, then the stationary diversity is approximated by a beta distribution. Since with Eq. (2.11), identical γ_k results in equivalent m_k and a_k, then the plausibility of Hubbell's neutral theory can be predicted by the same trade-off parameters for the quasi-neutral species, no matter how different their yields are. Conversely, in the

case with identical yields y_k but different trade-off parameters (γ_k, m_k, a_k), the selective term in Eq. (5.15) plays a large role in causing departure from the neutral results. Therefore, the life history trade-off plays an significant role in controlling the diversity of the local communities.

Above all, we answered the two central questions Alonso and McKane asked in [3] while generally accepting Hubbell's neutral theory as the null model for the community structure: " To what extent is the neutral assumption a good operational first approximation to describe ecological communities? Which mechanisms are responsible for any observed departures from neutrality? "

5.5.3 Mechanism of the Dispersal Difference

Dispersal-Limited Case

The last subsection compared Hubbell's neutral model with the neutrality defined at the individual level with our quasi-neutral demographic model with the neutrality defined at the species level, under the condition that the local community has low immigration rates. We found that Hubbell's neutral model is not sufficient to explain the mechanisms of the species difference when both have the largest fitness. When the dispersal is very limited (i.e., immigration rate is small), the quasi-neutral model shows that one species is superior to another, rather than Hubbell's supposition that the relative abundances are changed by chance and dispersal only.

Dispersal-Unlimited Case

We know that if the immigration rate m is close to 1 (dispersal-unlimited), there is a high probability that immigration will play a leading role in shaping the local community, with the result that the local abundance will be completely reflected in the abundance in the meta-community, i.e., $p_i = q_i$. Thus, we can envisage that the more connected the local-community is to the meta-community, the more the role of selection (advantages and disadvantages) in the local community will be weakened.

Furthermore, both in Figures 5.3 and 5.5, we found that when increasing the immigration rates, although there is no change in the mean relative abundance, the standard deviations of the distributions decrease greatly, which demonstrates a shrinking role of stochastic drift.

Therefore, difference in dispersal are an underlying mechanism that drives the composition and structure of the local communities.

5.6 Summary

In this chapter, we coupled the local chemostat community with a meta-community to determine how immigration influences the diversity of the quasi-neutral model. With a low immigration rate, the result of the strong selection does not change, the deterministic dynamics remains the same as the dynamic in Chapter 2. However, the stochastic model will no longer be trapped into the absorption state in the model with immigration present. Immigration plays a role as a mechanism to balance extinction and maintain a stationary diversity over an infinite time interval, even if its rate is very

small. Therefore, in a local community with very limited dispersal (low immigration rate), we can see that the deterministic selection may only explain the coexistence of the quasi-neutral species, and its stationary density abundance can be predicted through the stochastic drift, by assuming neutrality at the species level. This suggests that niche structure, stochastic drift and immigration all have key roles in shaping the structure of the chemostat community.

Later, the effects of the immigration rate and species intrinsic parameters on the stationary distribution were discussed. A small immigration rate which has no impact on the deterministic dynamics, plays a significant role in shaping the structure of the local community over infinite time interval. As the immigration rate increases, the roles played by stochastic drift and selection between quasi-neutral species are weakened. Since in the final diffusion approximation, the number of parameters is greatly reduced, we only need to discuss the effects of yields y_k and the sensitivity of the birth rate to the resource at the equilibrium state, γ_k. Although a simple advantage parameter could not be determined, there is still much information that may be collected. The yields y_k still do not play a large role if the trade-off parameter γ_k are identical, but γ_k becomes a selective disadvantage parameter when y_k is identical. That is to say that, in the model that the quasi-neutral species at the largest fitness have identical yields y_k but different trade-off parameters γ_k, the species with larger trade-off parameters has the advantage of dominating the community if immigration is absent, but has a reduced mean steady states relative abundance when immigration is introduced. If γ_k are identical for both quasi-neutral species, then the

application of Hubbell's Unified Theory of Biodiversity in our local community can be predicted, and all individual may be treated as ecologically equivalent, no matter how different their yields y_k are.

Due to the sensitivity of the species abundance to the parameters we observed, the weakness of Hubbell's neutral theory in shaping the structure of the local community with limited dispersal is found. Therefore, niche assembly together with Hubbell's neutral theory are not sufficient to explain the structure in a long time period in our chemostat model. Although it is still believed that emergent neutrality is reasonable with trade-off in complicated models, species differences amongst the quasi-neutral species could not be ignored in the dispersal assembly theory.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

In this thesis, by analysing competition among multiple species in a resource-limited chemostat environment, we determined how the structure of the local community is shaped, and investigated the role played by selection, stochastic drift and dispersal. To achieve this, a range of models were developed and compared: a deterministic model, a stochastic quasi-neutral model, a numerical model, as well as a stochastic quasi-neutral model with immigration and Hubbell's unified neutral model.

Firstly, we derived a niche assembly rule by developing a deterministic model in Chapter 2. By incorporating stochastic drift and limited dispersal in the first part of both Chapter 3 and Chapter 5, the deterministic limits of the dynamics will not change over compact time intervals in large communities. This gives a clear direction to the evolution, that is, the species sharing the same largest fitness in our model may survive and coexist. It

corresponds to the acquisition of the heritable adaptations to the current environment and allows species differentiation through life history trade-offs. Compared with the stationary distribution of the stochastic model in Chapter 5, the biodiversity given by the niche assembly rules merely predict the richness and a balance in the niches, where the coexistence equilibrium states are dependent on their initial frequencies and trade-offs parameters. In this case, at the beginning of time evolution, the structure of the communities is mainly regulated by the niche assembly rule, while the effects of stochastic drift and limited dispersal are not obvious.

Secondly, the mechanism of demographic drift was elaborated in Chapter 3 and Chapter 4, by including Markovian noise into our niche-structured quasi-neutral model. We found that, over infinite time intervals, the law of large numbers and central limit theory no longer holds in the case of large but still finite volume, rather the stochastic dynamics will be trapped into an absorbing state eventually due to stochastic drift. In order to analytically show how stochastic drift drives the population diversity to fixation and calculate more quantities, a diffusion approximation at a long time scale was derived. The development of the analytical method to derive this diffusion approximation is one of the most innovative parts of the thesis, whose details are presented in Sections 3.5 and 3.6. With the first attempt at the diffusion approximation, big discrepancies were observed when comparing with the simulation results in Chapter 4. To improve the accuracy of the analytical prediction, a correction was made on the derivatives of the projection map in Subsection 3.5.3.1, yielding a good match between the numerical and analytical results. The numerical model built in Chapter 4 regards the pop-

ulation dynamics as a discrete birth death random walk process. Although all the numerical results were obtained for the small populations due to the limitation of computational capacities, they are still well matched with the analytical prediction for large populations. Moreover, in the absence of immigration, the population in the local communities will be dominated by a single species with Gaussian distribution under the condition that extinction does not occur.

Dispersal is another key process in the local community that balances extinction to maintain the community diversity. In Chapter 6, by coupling the local chemostat community to a meta-community with limited dispersal, the stationary distribution of the relative abundance is explicitly calculated. Even with low immigration rates, the local diversity was sensitive to the abundance in the meta-community. By increasing the immigration rates, the role played by stochastic drift and selection between quasi-neutral species will be weakening.

Our quasi-neutral model is built to allow the species in the same niche to differ through life history trade-offs. Therefore, for all the models in this thesis, we discussed the effects of the trade-off parameter to determine whether the quasi-neutral species exhibits substantial differences by comparing with strictly neutral model. After that, the importance of the trade-off parameter γ_k which represents the sensitivity of growth rate to the resource concentration at the equilibrium state was demonstrated. γ_k helps to connect the fixation probabilities with the quasi-stationary distributions, and to predict a deviation from the strictly neutral model. This deviation could be ascribed to two separate phenomena: the first was an advantage in increasing chance

to dominate the population to the species with higher γ_k when immigration is absent in the model, the second was a disadvantage to this type on the final relative abundance. Furthermore, when the trade-off parameters are identical, the quasi-neutral species could be regarded as ecologically neutral, no matter how different the other parameter yields y_k are.

In conclusion, with sufficient quantitative results and clearly explanations, by thoroughly explaining the dynamics of the competitive species in the resource limited model, we show that the natural selection, stochastic drift and dispersal should be collectively instead of individually applied. The "niche assembly rules" which are a fitness consequence of the natural selection determine the richness, not the relative abundance, of the biodiversity. Over a long period of time, when the inferior species have already been screened out, niche assembly is no longer the dominant mechanism. At that moment, stochastic drift and dispersal together will play dominant roles to give the stationary abundance of the species in the local communities. The life history trade-off not only ensures the coexistence in the same niche, but also allows the quasi-neutral species to exhibit substantial difference by the effects of stochastic drift and dispersal.

6.2 Future Work and Directions

The strongest test of the quasi-neutral model would be to fit into experimental empirical data. However, to obtain real data is challenging. Microbial ecologists still are not sure how much data from a sample is sufficient to completely explain the biodiversity in a typical microbial community, and

then to test the model which we wish to extrapolate to a large population [63]. Also, the falsification of Hubbell's neutral model in a local community with low immigration rate does not mean rejection of neutrality. It is still the only sampling dispersal theory. Actually, Holt [29] finds the emergent property for neutrality in a macroscopic community: that when examined over certain spatial and temporal scales, a reasonable neutral approximation may be converged from non-neutral dynamical processes.

Therefore, although *a priori* theoretical models still need strong backing to support research into microbial communities, it is never useless to build more candidate models and iterate toward a predictive theoretical microbial ecology.

6.2.1 A Multivariate Quasi-Neutral Model

In this thesis, all our research focuses on a two-species quasi-neutral model, where the relative abundance of the first species $p(t)$ is sufficient to explain the behaviour of the whole population. The population process may be reduced to a one dimensional diffusion approximation. If we allow the number of species type $k > 2$, the relative abundance of each species is $\frac{\hat{x}_i}{\sum_j \hat{x}_j}$, and the fixation probability on the centre manifold is more than two dimensional. A multivariate quasi-neutral model could be derived for general applications in future.

6.2.2 A Dispersal-Assembly Model

Moreover, in developing the coupled model, the immigration rate is assumed to be small. Increasing the immigration rate, the deterministic dy-

namics will not be the same: the centre manifold will disappear, which may result in a different expression for the stationary distribution of the relative abundance. Further work on a general model for any immigration rates may be undertaken to discuss how the roles of the niche structure and dispersal difference will play out in shaping the structure of the community.

6.2.3 Numerical Test of Hubbell's Neutral Model

Although we have partly rejected the agreement between our niche structured quasi-neutral model and Hubbell's neutral model, the comparison is between two analytical diffusion approximations. Hubbell's numerical model is still the best model designed to apply to the samples[3]. Thus, the numerical model of Hubbell's unified theory of biodiversity on a chemostat model is worth testing in future work.

An approach to do this could involve three steps as in [31] [42]. First, estimate the two parameters: the fundamental biodiversity parameter θ and migration rate m. Second, generate 500 rank-abundance curves under the standard neutral model using the estimated parameters in the first step. Last step, test the hypothesis that the observed rank-abundance curve drawn from the distribution of simulated curves to our quasi-neutral stationary distribution.

6.2.4 A Spatial Model

All models in this thesis are build on the assumption that the chemostat is completely well stirred, so that the bacteria and nutrient are spatially uniformly distributed, and each individual has equal access to the nutrients.

In any real natural community, there will not only be one spatial scale, the spatial structure and its variance will greatly influences the ecological processes. Therefore an ideal theoretical attempt to explain the biodiversity of community requires an spatial formulation to connect the spatial pattern with the ecological process [3].

In future work, we could develop a stochastic spatial explicit model in spatially mixed local chemostat community, to analyse the relation between the spatial heterogeneity and the ecological resource.

To build the model, a discrete spatial structure is required. The local community will be comprised of a large number of components. The component labeled x has species density for the kth type $u(x,t)_k$ and resource concentration $S(x,t)$ at time t [1]. With the interaction among the components, a diffusion-reaction function may be present. Starting from the simple, deterministic, macroscopic description, successive layers of detail should be added into and discussed.

Appendix A: Simulation Results

Numerical Results in Figure 4.5

Table 1: Fixation Probability With Close γ_k

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$F(p\|\frac{y_1}{y_2}=1)$	0.086	0.196	0.282	0.393	0.489	0.60	0.698	0.786	0.890
$F(p\|\frac{y_1}{y_2}=0.1)$	0.115	0.185	0.29	0.405	0.51	0.605	0.675	0.82	0.91
$F(p\|\frac{y_1}{y_2}=10)$	0.09	0.2017	0.2995	0.3925	0.515	0.615	0.685	0.81	0.905

Numerical Results in Figure A of Figure 4.7 and Figure 4.8

Table 2: Fixation Probability With Identical y_k, but $\gamma_1 > \gamma_2$

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$F(p\|\frac{\gamma_1}{\gamma_2}=0.9625)$	0.086	0.196	0.282	0.393	0.489	0.60	0.698	0.786	0.890
$F(p\|\frac{\gamma_1}{\gamma_2}=1.3)$	0.11	0.23	0.34	0.46	0.548	0.654	0.749	0.825	0.934
$F(p\|\frac{\gamma_1}{\gamma_2}=2.425)$	0.196	0.392	0.482	0.59	0.67	0.758	0.818	0.898	0.934

Numerical Results in Figure B of Figures 4.7 and 4.8

Table 3: Fixation Probability With Identical y_k, but $\gamma_1 < \gamma_2$

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$F(p\|\frac{\gamma_1}{\gamma_2} = 0.9625)$	0.086	0.196	0.282	0.393	0.489	0.60	0.698	0.786	0.890
$F(p\|\frac{\gamma_1}{\gamma_2} = 0.5746)$	0.064	0.148	0.19	0.272	0.364	0.453	0.55	0.684	0.862
$F(p\|\frac{\gamma_1}{\gamma_2} = 0.2059)$	0.056	0.102	0.132	0.168	0.23	0.306	0.38	0.49	0.646

Numerical Results in Figure 4.9

Table 4: Mean of First Absorption Time with Identical y_k, But Different γ_k

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$ET(p\|\frac{\gamma_1}{\gamma_2} = 0.55)$	3807	6209	7505	8605	8821	8861	8109	7015	4451
$ET(p\|\frac{\gamma_1}{\gamma_2} = 1)$	3911	6603	8109	9271	9064	8473	7622	6829	4446
$ET(p\|\frac{\gamma_1}{\gamma_2} = 1.5)$	3998	6698	8194	8974	9276	8986	8058	6252	4232
$ET(p\|\frac{\gamma_1}{\gamma_2} = 3)$	4287	6917	7769	8119	8457	8021	7424	5963	3808

Bibliography

[1] Andrews, S.S., Dinh, T. and Arkin, A.P. :*Stochastic Models of Biological Processes* Encyclopaedia of Complexity and System Science, Vol 9, Springer (2009), pp.8730-8749.

[2] Allouche, O. and Kadmon, R. :*Demographic Analysis Of Hubbell's Neutral Theory Of Biodiversity.* Journal of Theoretical Biology 258 (2009), pp.274-280.

[3] Alonso, D., Etienne, R. S. and McKane, A. J : *The Merits of Neutral Theory* Trends in Ecology and Evolution Vol. 21 No. 8 , (2006), pp.451-457

[4] Artalejo, J. R., and Lopez-Herrero, M. J. :*Quasi-Stationary and Ratio of Expectations Distribution: A Comparative Study* Journal of Theoretical Biology, 266 (2010), pp.264-274

[5] Bacaër, N : *A Short History of Mathematical Population Dynamics.* Springer (2011).

[6] Bell, G : *Neutral Macro-ecology.* Science 293, (2001), 2413

[7] Burslem, D., Pinard, M. and Hartley, S : *Biotic Interactions in the Tropics: Their Role in the Maintenance of Species Diversity.* Cambridge University Press pp.107-138

[8] Campillo, F., Joannides, M. and Larramendy-Valverde, I :*Approximation Of The Fokker-Planck Equation of the Stochastic Chemostat.* arXiv:1111.5716v1 [math.PR] 24 Nov (2011)

[9] Campillo, F., Joannides, M. and Larramendy-Valverde, I :*Stochastic Model of the Chemostat.* RINRIA, Nov (2010)

[10] Collet, P., Martinez, S., Meleard, S. and Martin, J. S. :*Stochastic Models for a Chemostat and Long Time Behaviour.* arXiv:1206.3691v1 [math.PR] June 19, (2009)

[11] Darwin, C. : *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* London: John Murray, Albemarle Street. (1859)

[12] Durrett, R. and Levin, S : *The Importance of Being Discrete and Spatial* Theory. Popul. Biol. 46 (1994), pp. 363-394

[13] Ethier, S. N. and Kurtz, T. G. : *Markov Processes: Characterization and Convergence.* John Wiley and Sons, New York, (1986)

[14] Euler, L. :*Introductio in analysin infinitorum* Tomus primus. Bousquet, Lausanne, (1798)

[15] Ewens, W. J. : *Mathematical Population Genetics: Theoretical Introduction*

[16] Ewens, W. J. : *Mathematical Population Genetics* Springer (2004)

[17] Fargione, J., Brown, C. S., and Tilman, D : *Community Assembly and Invasion: An Experimental Test of Neutral Versus Niche Processes* PNAS Vol. 100, No. 15, July (2003)

[18] Farkas, Z. and Fulop, T : *One Dimensional Drift-diffusion Between Two Absorbing Boundaries: Application to Granular Segregation* Journal of Physics A: Mathematical and General, Vol. 34, No.15, Oct (2000)

[19] Fierer, N and Lennon, J. T. :*The Generation And Maintenance of Diversity in Microbial Communities.* American Journal of Botany Vol. 98, No. 3, (2011), pp. 439-448

[20] Fisher, R. A. :*The Genetical Theory of Natural Selection.* Clarendon Press, Oxford(1930). www.archive.org

[21] Fox, J. W. and Srivastava, D :*Predicting Local-Regional Richness Relationships Using Island Biogeography Models* Oikos 113, pp. 376-382.

[22] Gandon, S., Lambert, A. and Parsons, T. : *Stochastic Contributions To The Evolution Of Two-Strain SIR Epidemics.* May 18, (2011)

[23] Gillespie, D, T. :*Exact Stochastic Simulation of Coupled Chemical Reactions* J.Phys. Chem., Vol.81, No. 25, (1977), pp.2340-2361

[24] Glendinning, P : *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations.* Cambridge University Press (1994)

[25] Grimm, V. and Wissel, C. : *The Intrinsic Mean Time to Extinction: A Unifying Approach to Analysing Persistence and Viability of Populations.* OIKOS, Vol. 105, No. 3, (2004)

[26] Hansen, S. R. and Hubbell, S. P. : *Single-Nutrient Microbial Competition: Qualitative Agreement between Experimental and Theoretically Forecast Outcomes.* Science, New Series, Vol. 207, No. 4438, Mar, (1980), pp. 1491-1493

[27] Hardy, G. H. :*Mendelian Proportions in a Mixed Population.* Science 28, (1908), pp. 49-50

[28] Herbert, D., Elsworth, R. and Telling, R. C. : *The Continuous Culture of Bacteria: a Theoretical and Experimental Study.* J.gen.Microbial. 14, (1956), pp. 601-622

[29] Holt, R. D. :*Emergent Neutrality.* Trends in Ecology and Evolotion Vol. 21, No. 10, Aug (2006), pp. 531-533

[30] Houchmandzadeh, B. and Vallade, M. :*Clustering In Neutral Ecology.* Physical Review E 68, 061912 (2003)

[31] Hubbell, S. P. : *The Unified Neutral Theory of Biodiversity and Biogeography.* Princeton University Press (2001).

[32] Hubbell, S. P. : *Neutral Theory and The Evolution of Ecological Equivalence.* Ecology, Vol. 87, No. 6, (2006), pp. 1387-1398

[33] Hsu, S. B., Hubbell, S. and Waltman, P. : *A Mathematical Theory for Single-nutrient Competition in Continuous Cultures of Microorganisms.* Appl. Math. Vol. 32, No. 2, Mar (1977)

[34] Kampen, V. : *Stochastic Processes in Physics and Chemistry* Elsevier, Amsterdam (1981)

[35] Kimura, M : *On the Probability of Fixation of Mutant Genes in a Population.* Genetics, Vol. 47, No. 6, June (1962), pp. 713-719

[36] Krone, S. M. :*Spatial Models: Stochastic and Deterministic.* Mathematical and Computer Modelling, 40 (2004), pp. 393-409

[37] Kurtz, T.s G. and Protter, P. : *Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations.* The Annals of Probability, Vol. 19, No. 3, (1991) pp. 1035-1070

[38] Laska, M. S. and Wootton, J. T. :*Theoretical Concepts and Empirical Approaches to Measuring Interaction Strength* Ecology, Vol. 79, No. 2, (1998), pp. 461-476

[39] Lokta, A. J. :*Analytical Note on Certain Rhythmic Relations in Organic System* Proc. Natl. Acad. Sci. Vol.6, (1920), pp. 410-415

[40] Makarov, Y. V. and Dong, Z. Y. : *Eigenvalues and Eigenfunctions*, Jun (2000)

[41] Malthus, T. : *An Essay on the Principle of Population.* London (1798).

[42] McGill, B. J. : *A Test of the Unified Neutral Theory of Biodiversity.* Nature, Vol. 422, No.24, April (2003)

[43] McGill,B. J., Maurer, B. A. and Weiser, M. D. :*Empirical Evaluation of Neutral Theory.* Ecollgy, Vol. 87, No. 6, (2006), pp. 1411-1423

[44] McKane, A. J., Alonso, D. and Sole, R. V. : *Analytic Solution of Hubbell's Model of Local Community Dynamics.* Theoretical Population Biology Vol. 65, (2004), pp. 67-73

[45] Meucci, A: *Review of Statistical Arbitrage, Cointegration, and Multivariate Ornstein-Uhlenbeck.*

[46] Moran,P. A. P. : *Random Processes in Genetics.* Mathematical Proceedings of the Cambridge Philosophical Society, Vol. 54, (1958), pp. 60-71

[47] Mouquet, N. and Loreau, M : *Community Patterns in Source-Sink Meta-communities.* American Naturalist, Vol. 162, (2003), pp.544-557

[48] Ovaskainen, O and Meerson, B : *Stochastic Models Of Population Extinction* Trends in Ecology And Evolution, Vol. 25, Issue 11, Nov (2010), pp.643-652

[49] Oliveira,M. M. and Dickman, R. : *How to Simulate The Quasi-stationary State* Physical Review E 71,016129 (2005)

[50] Parsons, T : *Quasi-Neutrality in An Explicit Resource Model.*

[51] Parsons, T : *Limit Theorem for Competitive Density-denpendent Population Processes.*

[52] Parsons, T. L., Quince, C. and Plotkin, J. B. : *Some Consequences of Demographic Stochasticity in Population Genetics.* Genetics 185 Aug (2010), pp.1345-1354

[53] Parsons, T. L., Quince, C. : *Fixation in Haploid Populations Exhibiting Density Dependence: The Non-neutral Case.* Theoretical Population Biology 72 (2007), pp. 121-135

[54] Parsons, T. L. and Quince, C. : *Fixation in Haploid Populations Exhibiting Density Dependence: The Quasi-neutral Case.* Theoretical Population Biology 72 (2007), pp. 468-479

[55] Riedler, M. G. and Buckwar, E : *Law of Large Numbers and Langevin Approximations for Stochastic Neural Field Equations.* Journal of Mathematical Neuroscience Vol. 3, No. 1, (2013)

[56] Roberts, A. J. : *Normal Form Transforms Separate Slow and Fast Modes in Stochastic Dynamical System.* Physica A Vol. 387, No. 12, (2008).

[57] Paulo, S. and Roos, P. :*Fast and Deterministic Computation of Fixation Probability in Evolutionary Graphs.* The Sixth IASTED Conference on Computational Intelligence and Bioinformatics (2011)

[58] Pfaffelhuber, P., Pennings, P., and Hermisson J. :*Population Genetic Tutorial: Chapter 2, The Wright-Fisher model and the Neutral Theory* University of Vienna, February 2, 2009

[59] Purves, D. W. and Pacala, S. W. :*Ecological Drift in Niche-structured Communities: Neutral Pattern Does not Imply Neutral Process.* Biotic Interactions in the Tropics Their Role in the Maintenance of Species Diversity (Edited by Burslem, D., Pinard, M. , Hartley, S.), Cambridge University Press, pp. 107-138

[60] Purves, D. W. and Turnbull, L. A. :*Different But Equal: The Implausible Assumption at the Heart of Neutral Theory* Journal of Animal Ecology 79 (2010), pp. 1215-1255

[61] Silchenko, A.N., Luchinsky, D.G. and McClintock, P.V.E : *Noise-Induced Escape Through a Fractal Basin Boundary.* Physica A 327 (2003), pp. 371-377

[62] Sloan, W. T., Lunn, M. and Curtis, T. P. : *Quantifying the Roles of Immigration and Chance in Shaping Prokaryote Community Structure.* Environmental Microbiology Vol. 8, No. 4, (2006), pp. 732-740

[63] Sloan, W. T., Woodcock, S., Lunn, M., Head, I. M. and Curtis, T. P. :*Modeling Taxa-Abundance Distributions in Microbial Communities using Environmental Sequence Data* Microbial Ecology, Vol. 53, No. 3, Apr (2007) pp. 443-455

[64] Tilman, D. : *Niche Tradeoffs, Neutrality, and Community Structure: A Stochastic Theory of Resource Competition, Invasion, and Community Assembly.* PNAS Vol. 101, No. 30 July (2004)

[65] Tilman, D. :*Competition and Biodiversity in Spatially Structured Habitats.* Ecology, 75, (1994), pp. 2-16

[66] Vallade, M, and Houchmandzadeh, B : *Analytical Solution of a Neutral Model Of Biodiversity.* Physical Review E. 68, 061902 (2003)

[67] Vellend, M. :*Conceptual Synthesis in Community Ecolgy* The Quarterly Review of Biology, Vol. 85, No. 2 June (2010), pp. 183-206

[68] Verhulst, P. F. : *Notice Sur La Loi Que La Population Poursuit Dans Son Accroisse- ment.* Corresp. Math. Phys. 10, (1838), pp. 113-121

[69] Volkov, I., Banavar, J. R., Hubbell, S. P. and Maritan, A :*Neutral Theory and Relative Species Abundance In Ecology* Nature, Vol. 424, No. 28, Aug (2001)

[70] Volterra, V. :*Variations and fluctuations of the Number of Individuals in Animal Species Living Together in Animal Ecology.* Chapman, McGrawHill, (1931)

[71] Walker, S. C. and Cyr, H. : *Testing The Standard Neutral Model of Biodiversity in Lake Communities.* Oikos 116, (2007), pp. 143-155

[72] Watson, H. W., and Galton, F. :*On the Probability of the Extinction of Families.* The Journal of the Anthropological Institute of Great Britain and Ireland, Vol. 4, (1875), pp. 138-144.

[73] Wakeley, J.:*Coalescent Theory: An Introduction.* Roberts Company Publishers, Greenwood Village, Colorado. (2008)

[74] Whitman, W. B., Coleman, D. C. and Wiebe, W. J. :*Prokaryotes: The Unseen Majority.* Proceedings of the National Academy of Sciences of the United States of America 95, pp. 6578-6583

[75] Woodcock, S., Gast, C. J., Bell, T., Lunn, M., Curtis, T. P., Head, I. M. and Sloan, W.T. : *Neutral Assembly of Bacterial Communities.* FEMS Microbiol Ecol, 62 (2007), pp. 171-180

[76] Weinberg, W. *Uberden Nachweisder Vererbung Beim Menschen.* Jahreshefte Des Vereins Fur Vaterlandische Naturkunde in Wurttemberg 64, (1908). pp. 368-382.

[77] Wright, S. :*Evolution in Mendelian Populations.* Genetics 16, (1931), pp. 97-159

[78] Wright, S. :*Evolution and the Genetics of Populations, Vol. 2, Theory of Gene Frequencies.* University of Chicago Press, (1969)

[79] *Lecture Notes of Quantitative Biology: 6.3 Microbial Growth in a Chemostat.* University of Warwick