Spatial populations with seed-bank

Oomen, M.

Introduction

§1.1 Spatial populations with seedbank

In populations with a *seed-bank*, individuals can temporarily become dormant and re-frain from reproduction, until they can become active again. Seed-banks are observed in many taxa, including plants, bacteria and other micro-organisms. Typically, they arise as a response to unfavourable environmental conditions. The dormant state of an individual is characterised by low metabolic activity and interruption of phenotypic development (see e.g. [55]). After a varying and possibly large number of generations, a dormant individual can be resuscitated under more favourable conditions and re-prise reproduction after having become active again. This strategy is known to have important implications for population persistence, maintenance of genetic variability and stability of ecosystems, (see e.g. [54]). It acts as a *buffer* against evolutionary forces such as genetic drift, selection and environmental variability.

Various attempts were made to include a seed-bank in already existing mathem-atical models that describe the genetic evolution of populations (see [50], [11], [10] and [70].) However, after inclusion of the seed-bank these models become complex, because they have long memory. In [12] the so-called "two-type Fisher-Wright model with seed-bank" was introduced. This was the first model that describes the evolution of a population with seed-bank as a Markov process. In this model individuals move in and out of the seed-bank at prescribed rates. Outside the seed-bank individuals are subject to *resampling*, while inside the seed-bank their resampling is *suspended*. Both the long-time behaviour and the genealogy of the population were analysed in detail. In particular, it was shown that the seed-bank increases the genetic variability of the population.

The goal of this thesis is to extend the seed-bank model introduced in [12] to the *spatial* setting where individuals can migrate between different colonies. We analyse the long-time behaviour of the evolution of a spatial population with seed-bank in different settings. We show how the seed-bank increases the genetic variability, compared to spatial population models without seed-bank. In particular, we show how certain types of seed-banks can even prevent the loss of genetic variability altogether.

§1.2   Modeling population genetics

We give a short introduction to modeling genetic evolution of populations. We also introduce two important tools to analyse the genetic evolution in populations, namely, the *Kingman coalescent* and *duality*.

§1.2.1   The Fisher-Wright model

One of the driving forces in the genetic evolution of populations is genetic drift. Genetic drift is the evolutionary mechanism that selects genes randomly. To illustrate the concept of genetic drift, consider a population of turtles. Each year the turtles lay plenty of eggs on the beach, but only a few of these eggs grow into a mature turtle. Which of the eggs will do so is random. In this way, randomness plays an important part in the genetic evolution of populations, and this randomness is called *genetic drift*.

In mathematics genetic drift is modeled through the *Fisher-Wright* model. In the Fisher-Wright model we consider a population of N individuals. Each individual can carry one of two genetic types, denoted by $\heartsuit$ and $\diamondsuit$. In each generation all the individuals will reproduce themselves according to the following rule:

- Each individual chooses uniformly at random an individual from the population and adopts its type. The chosen individual may be the individual itself.

This way of modeling reproduction is called *resampling*. Note that the number of individuals remains fixed during the evolution. Since we mostly consider very large populations without external evolutionary forces, we may assume the population size to be approximately constant. Therefore the assumption that the population size remains fixed is reasonable. An example of a population of 5 individuals is shown in Fig. 1.1. The resampling mechanism can be interpreted as follows: Each individual in the population gets a random number of offspring between 1 and N, and the total number of offspring in the next generation is N. This gives a more natural interpretation of resampling, but the way resampling is phrased above makes the mathematical analysis easier.

Evolution equation.   To describe the genetic evolution in the population, we analyse the fraction of individuals of type $\heartsuit$. Label the N individuals by $[N] = \{1, \ldots, N\}$. Define $\xi(k) = (\xi_j(k))_{j \in [N]} \in \{0,1\}^{[N]}$ as the random vector where $\xi_j(k) = 1$ if the j'th individual is of type $\heartsuit$ at time k and $\xi_j(k) = 0$ if the j'th individual is of type $\diamondsuit$ at time k. Then

$$X^N(k) = \frac{1}{N} \sum_{j \in [N]} 1_{\{\xi_j(k) = 1\}} \tag{1.1}$$

is the fraction of individuals of type $\heartsuit$ in generation k. Since there are two types of individuals in the population, the fraction of individuals of type $\diamondsuit$ in generation k is given by $1 - X^N(k)$. The distribution of $X^N(k+1)$ given $X^N(k)$ is $\mathrm{BIN}(N, X^N(k))$. A key question is whether eventually there is only one type of individual left in the
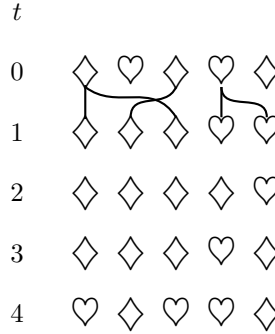
t

Figure 1.1: Example of the evolution for a population with $N = 5$ individuals in 5 generations. The solid lines within the active population represent resampling in the first generation.

population, or there are always two types of individuals in the population. If only one of the two types is left, then we say that *genetic variability* is lost. For a finite population, genetic variability is lost eventually. The expected time until this happens is of order N, the size of the population [30].

In this thesis we look at genetic evolution in populations where the number of individuals tends to infinity. This large population model is referred to as the continuum model. To obtain this continuum model, we let the number of individuals in the population tend to infinity and speed up time proportionally to the number of individuals in the population, i.e.,

$$\lim_{N \to \infty} \mathcal{L}[\left(X^N(\lfloor Nt \rfloor)\right)_{t \geq 0}]. \tag{1.2}$$

Thus, we observe larger and larger populations on time scales where these populations start to loose their genetic variability. The limit in (1.2) is the law of the continuous-time process $(x(t))_{t \geq 0}$ that evolves according to the stochastic differential equation (SDE)

$$\mathrm{d}x(t) = \sqrt{x(t)(1 - x(t))}\,\mathrm{d}w(t), \tag{1.3}$$

where $(w(t))_{t \geq 0}$ is a standard Brownian motion. Its initial law is given by $\mathcal{L}[x(0)] = \lim_{N \to \infty} \mathcal{L}[X^N(0)]$. The process $(x(t))_{t \geq 0}$ evolving according to (1.3) is called the *Fisher-Wright diffusion* and has state space $[0, 1]$. The stochastic differential equation (SDE) in (1.3) has a unique solution that is a Markov process (see [72]). The fixed points of (1.3) are 0 (only individuals of type $\diamondsuit$ are left) and 1 (only individuals of type $\heartsuit$ are left). The Fisher-Wright diffusion reaches its fixed points in finite time [30].

Geneology. After sampling $n \in \mathbb{N}$ individuals from a large population at some large time, we can ask ourselves what the lineages of these n individuals are. If two of the n sampled individuals have a common ancestor a time s backwards in time, then their lineages coalesce, (see Figure 1.2). It turns out that if we sample individuals from the continuum model, then any two lineages coalesce at rate 1, independently
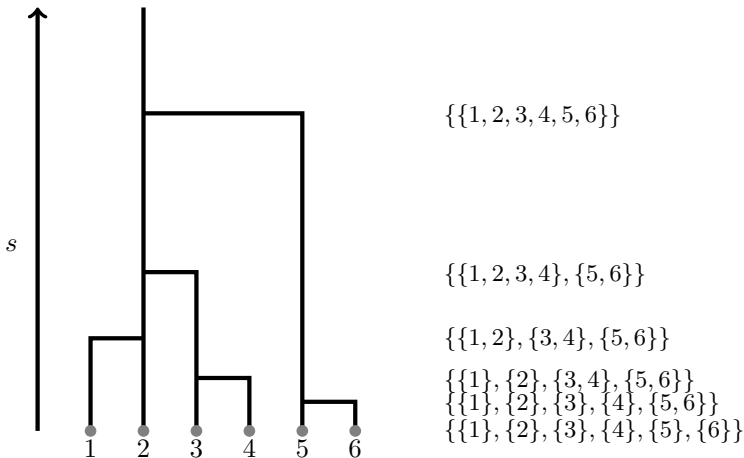
Figure 1.2: Example of a genealogy of 6 individuals sampled from a Fisher-Wright diffusion. The corresponding Kingman coalescent is written on the right. Each time two lineages merge, the corresponding partition elements merge. Time is indicated by s and is running backwards.

of the other lineages. The ancestral lineages together are called the *genealogy* of the n individuals.

The process that formally describes the genealogy of the n sampled individuals in the continuum model is called the *Kingman coalescent*. The Kingman coalescent is a partition-valued process that at time $s = 0$ assigns to each of the n individuals a partition element, i.e., at time $s = 0$ the Kingman coalescent starts from state $\{\{1\}, \{2\}, \ldots, \{n\}\}$. If two lineages coalesce, then the corresponding two partition elements of the Kingman coalescent merge (see Fig. 1.2). Thus, any two partition elements merge at rate 1, independently of the other partition elements. The Kingman coalescent describes how the genetic evolution of a population took place in the past. The Kingman coalescent runs backwards in time. For this reason it is sometimes called *the backward process*. In contrast, the Fisher-Wright diffusion is called *the forward process*.

Since in the Fisher-Wright model individuals inherit their type from their parents, any two individuals whose lineages have a common ancestor are of the same type. If the number of sampled individuals tends to infinity, then the related ancestral lineages still have a common ancestor a finite time s backwards, (see [30]). Consequently, all the individuals in the population are of the same type and genetic variability is lost. We say that *the Kingman coalescent comes down from infinity in finite time*. This is the backward counterpart of the fact that the Fisher-Wright diffusion hits its fixed points in finite time.

Duality. Related to the coalescent process is the *block-counting process*. Suppose that at some large time $t > 0$ we sample n individuals from a population evolving ac-

cording to the Fisher-Wright diffusion in (1.3). The block counting process $(N(s))_{s\geq 0}$ counts the number of ancestral lineages when we traverse s backwards in time,

$$N(s) = \# \text{ lineages left at time } s. \tag{1.4}$$

Since any two lineages merge at rate 1, the process $(N(s))_{s\geq 0}$ has transition rates

$$n \to n-1 \text{ at rate } \binom{n}{2}. \tag{1.5}$$

Therefore the block-counting process is a death-process. Like the Kingman coalescent, the block-counting process is a backward process.

Let $(x(t))_{t\geq 0}$ be the Fisher-Wright diffusion starting from state $x \in [0,1]$. Let $(N(t))_{t\geq 0}$ be the block-counting process starting from $n \in \mathbb{N}$. Then, for all $n \in \mathbb{N}$, $x \in [0,1]$ and $t \geq 0$, the following relation holds:

$$\mathbb{E}^x[(x(t))^n] = \mathbb{E}^n[x^{N(t)}]. \tag{1.6}$$

Here the expectation on the left-hand side is taken over the Fisher-Wright diffusion $(x(t))_{t\geq 0}$ and the expectation on the right-hand side is taken over the block-counting process $(N(t))_{t\geq 0}$. The relation is called *moment duality*. This moment duality allows us to calculate all the moments of the Fisher-Wright diffusion at a given time t in terms of the death process at time t, which is simple to analyse. Note that the duality relation also expresses a relation between the backward and the forward processes.

State-dependent resampling rates. In the Fisher-Wright model individuals resample at rate 1. However, it is natural to allow for resampling rates that depend on the state of the population. To do this, let $g : [0,1] \to \mathbb{R}_{\geq 0}$ be any function that satisfies

- $g(0) = g(1) = 0$,

- $g(x) > 0$ for $x \in (0,1)$,

- g is Lipschitz continuous on $[0,1]$.

The evolution of the continuum model with resampling function g is given by

$$\mathrm{d}x(t) = \sqrt{g(x(t))}\,\mathrm{d}w(t), \tag{1.7}$$

where $(w(t))_{t\geq 0}$ is a standard Brownian motion. If we choose $g(x) = x(1-x)$, we recognize the Fisher-Wright diffusion in (1.3). Since the Fisher-Wright diffusion resamples at rate 1, the resampling rate in state x for the model in (1.7) is $\frac{g(x)}{x(1-x)}$. The first condition on g ensures that once the genetic diversity is lost, i.e., there are only $\heartsuit$ or $\diamondsuit$ left in the population, it cannot return.

The drawback of the continuum model in (1.7) is that it does not have a duality relation as in (1.6). Therefore this model is more difficult to analyse. Comparing the continuum model in (1.7) with the continuum models where $g = dx(1-x)$ for some constant $d \in (0,\infty)$, we are still able to analyse (1.7). This technique is called *comparison*.

Extensions of the Fisher-Wright model. The Fisher-Wright model can be extended in several ways to include other evolutionary forces. For example, selection of a fitter type and mutation of genes can be included. Also, more than two gene types can be included, which leads to the *multi-type Fisher-Wright model*. The extension to infinitely many gene types is called the *Fleming Viot model*. In the spatial Fisher-Wright model, there are multiple colonies, each evolving according to the Fisher-Wright model, and individuals are allowed to migrate. For all these extensions, extensive research was done. For an overview of the state of the art we refer the reader to [4]. The addition of a seed-bank to the Fisher-Wright model is relatively new and was introduced in 2016 in [12]. The Fisher-Wright model with seed-bank will be the building block of the spatial models considered in this thesis, and is introduced in the next section.

§1.2.2 The Fisher-Wright model with seed-bank

The Fisher-Wright model with (strong) seed-bank defined in [12] consists of a *single colony* with $N \in \mathbb{N}$ active individuals and $M \in \mathbb{N}$ dormant individuals. Each individual can carry one of two types: $\heartsuit$ or $\diamondsuit$. Let $\epsilon \in [0,1]$ be such that ϵN is integer and $\epsilon N \leq M$. Put $\delta = \frac{\epsilon N}{M}$. The evolution of the population is described by a discrete-time Markov chain that undergoes four transitions per step:

(1) From the N active individuals, $(1-\epsilon)N$ are selected uniformly at random without replacement. Each of these selected individuals resamples, i.e., adopts the type of an active individual selected uniformly at random with replacement, and remains active.

(2) Each of the ϵN active individuals not selected resamples, i.e., adopts the type of an active individual selected uniformly at random with replacement, and subsequently becomes dormant.

(3) From the M dormant individuals, $\delta M = \epsilon N$ are selected uniformly at random without replacement. These selected individuals become active. Since these individuals come from the dormant population, they do not resample.

(4) Each of $(1-\delta)M$ dormant individuals not selected remains dormant and retains its type.

Note that the total sizes of the active and the dormant population remain fixed. During the evolution the dormant and the active population *exchange* individuals. We will refer to the repository of the dormant population as the seed-bank. Fig. 1.3 depicts the first five generations of a population with 5 individuals in the active population and 3 individuals in the dormant population. Fig. 1.3 also shows how in the Fisher-Wright model with seed-bank genetic variability in the active population can be lost, but can be reintroduced again due to the exchange with the dormant population.
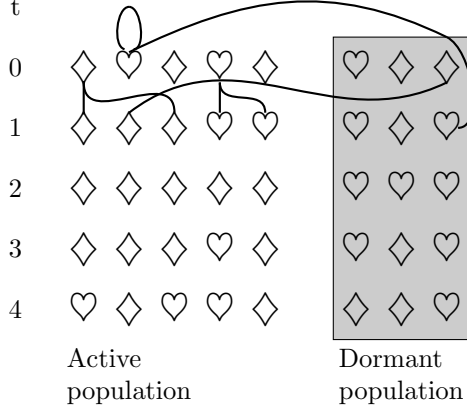
Figure 1.3: Example of the evolution for a population with $N = 5$ active individuals and $M = 3$ dormant individuals. The solid lines within the active population represent resampling, those between the active and the dormant population represent exchange with the seed-bank. Only 1 active individual and 1 dormant individual exchange places per unit of time, which corresponds to $\epsilon = \frac{1}{5}$ and $\delta = \frac{1}{3}$. The relative size of the dormant and the active population is $K = \frac{3}{5}$. Note that the genetic diversity in the active population is lost in generation $t = 2$, but returns in generation $t = 3$ via the seed-bank.

Evolution equation. To formally describe the Fisher-Wright model with seed-bank, we keep track of the fractions of individuals of type $\heartsuit$ in the active and the dormant population. Let $c = \epsilon N = \delta M$, i.e., c is the number of pairs of individuals that change state. Label the N active individuals from 1 to N and the M dormant individuals from 1 up to M. Write $[N] = \{1, \ldots, N\}$ and $[M] = \{1, \ldots, M\}$. Let $\xi(k) = (\xi_j(k))_{j \in [N]} \in \{0,1\}^{[N]}$ be the random vector where $\xi_j(k) = 1$ if the j'th active individual is of type $\heartsuit$ at time k and $\xi_j(k) = 0$ if the j'th active individual is of type $\diamondsuit$ at time k. Similarly, we let $\eta(k) = (\eta_j(k))_{j \in [M]} \in \{0,1\}^{[M]}$ be the random vector where $\eta_j(k) = 1$ if the j'th dormant individual is of type $\heartsuit$ at time k and $\eta_j(k) = 0$ if the j'th dormant individual is of type $\diamondsuit$ at time k. Define

$$X^N(k) = \frac{1}{N} \sum_{j \in [N]} \mathbf{1}_{\{\xi_j(k)=1\}},$$

$$Y^M(k) = \frac{1}{M} \sum_{j \in [M]} \mathbf{1}_{\{\eta_j(k)=1\}}. \tag{1.8}$$

Like for the Fisher-Wright model without seed-bank, we can pass to the continuum model. To do so we let both the active and the dormant population size tend to infinity, while keeping their relative sizes fixed, and speed up time proportional by the size of the active population, i.e.,

$$\lim_{N \to \infty} \mathcal{L}\left[\left(X^N(\lfloor Nt \rfloor), Y^M(\lfloor Nt \rfloor)\right)_{t \geq 0}\right] = \mathcal{L}\left[(x(t), y(t))_{t \geq 0}\right]. \tag{1.9}$$

Define

$$K = \frac{\text{size dormant population}}{\text{size active population}} = \frac{M}{N}, \tag{1.10}$$

which is the relative size of the dormant population compared to the active population. It was shown in [12] that the limiting process $(x(t), y(t))_{t\geq 0}$ in (1.9) evolves according to the stochastic differential equation

$$\begin{aligned}
\mathrm{d}x(t) &= c\left[y(t) - x(t)\right] \mathrm{d}t + \sqrt{x(t)(1 - x(t))}\, \mathrm{d}w(t), \\
\mathrm{d}y(t) &= \frac{c}{K}\left[x(t) - y(t)\right] \mathrm{d}t,
\end{aligned} \tag{1.11}$$

where $(w(t))_{t\geq 0}$ is a standard Brownian motion. The first term in the first line of (1.11) and the term in the second line of (1.11) describe the exchange of active and dormant individuals in the population. The second term in the first line of (1.11) describes the resampling in the active population. Note that the dormant population does not resample and hence evolves only due the exchange with the active population. In [12] it was shown that in the continuum Fisher-Wright model with seed-bank eventually only one type is left.

For later generalisations, we define the exchange rate

$$e = \frac{c}{K}, \tag{1.12}$$

and rewrite equation (1.11) as

$$\begin{aligned}
\mathrm{d}x(t) &= Ke\left[y(t) - x(t)\right] \mathrm{d}t + \sqrt{x(t)(1 - x(t))}\, \mathrm{d}w(t), \\
\mathrm{d}y(t) &= e\left[x(t) - y(t)\right] \mathrm{d}t.
\end{aligned} \tag{1.13}$$

The continuum process $(x(t), y(t))_{t\geq 0}$ evolving according to (1.13) will be the building block of the models analyzed in this thesis. The evolution given in (1.13) is schematically depicted in Fig. 1.4.

The seed-bank coalescent. To describe the genealogy of the continuum Fisher-Wright model with seed-bank, we sample n active individuals and m dormant individuals from the population at some large time and describe their ancestral lineages. We distinguish between active and dormant lineages by giving them labels A for active and D for dormant. The lineages behave according to the following rules:

- Each pair of active lineages coalesces at rate 1, independently of all other lineages.

- Each active lineage becomes dormant at rate Ke.

- Each dormant lineage becomes active at rate e.

Note that dormant lineages cannot coalesce: they can only become active.

Formally the lineages are decribed by the so-called *seed-bank coalescent* that was introduced in [12]. Like the Kingman coalescent, the seed-bank coalescent is
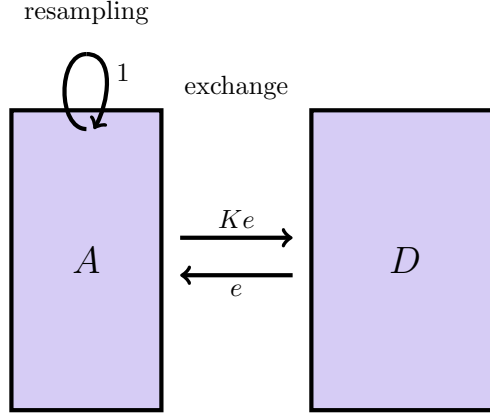
resampling

1

exchange

Ke

e

A

D

Figure 1.4: Schematic picture of the Fisher-Wright diffusion with seed-bank in (1.13). Active individuals resample at rate 1 and become dormant at rate Ke. Dormant individuals become active at rate e.

a partition-valued process, but in the seed-bank coalescent each partition element is labeled by A or D to indicate whether the corresponding lineage is active or dormant. In [12] it was shown that, as n and m tend to infinity, the ancestral lineages no longer have a common ancestor a finite time back. Hence *the seed-bank coalescent does not come down from infinity in finite time*. This result shows that the Fisher-Wright model with seed-bank behaves qualitatively differently than the Fisher-Wright model seed-bank.

Duality for the seed-bank model. For the Fisher-Wright model with seed-bank we have a similar duality relation as for the Fisher-Wright model without seed-bank. Let

$$N(s) = \# \text{ active lineages left at time } s,$$
$$M(s) = \# \text{ dormant lineages left at time } s. \tag{1.14}$$

Then the block-counting process $(N(s), M(s))_{s\geq 0}$ has transition rates

$$(n, m) \rightarrow \begin{cases} (n-1, m), & \text{at rate } \binom{n}{2}, \\ (n-1, m+1), & \text{at rate } nKe, \\ (n+1, m-1), & \text{at rate } me. \end{cases} \tag{1.15}$$

Let $(x(t), y(t))_{t\geq 0}$ be the Fisher-Wright diffusion with seed-bank evolving according to (1.13) and starting from state $(x, y) \in [0,1]^2$. Let $(N(t), M(t))_{t\geq 0}$ be the block-counting process starting from state $(n, m) \in \mathbb{N}^2$. Then, for all $(n, m) \in \mathbb{N}^2$, $(x, y) \in [0,1]^2$ and $t \geq 0$, the following relation holds:

$$\mathbb{E}^{(x,y)}[(x(t))^n (y(t))^m] = \mathbb{E}^{(n,m)}[x^{N(t)} y^{M(t)}]. \tag{1.16}$$

Here, the expectation on the left-hand side is taken over the Fisher-Wright diffusion with seed-bank, and the expectation on the right-hand side is taken over the block-

counting process starting in (n, m). Thus, also the Fisher-Wright model with seed-bank has a *moment dual*.

Wake-up time distribution of individuals. It has been recognised that qualitatively different behaviour may occur when the wake-up time of individuals in the seed-bank changes from having a thin tail to having a fat tail [55], [50]. Fat-tailed behaviour of the wake-up times is observed in colonies of bacteria. The drawback of the Fisher-Wright model with seed-bank is that it gives thin tails for the wake-up time of individuals. If we define the wake-up time

$$\tau = \text{ time a lineage spends in the seed-bank before it wakes up again}, \qquad (1.17)$$

then

$$\tau \overset{d}{=} \text{EXP}(e). \qquad (1.18)$$

In Section 1.2.3 we will show how we can adapt the Fisher-Wright model with seed-bank to allow for more general wake-up times *without loosing the Markov property*.

Extensions of the Fisher-Wright model with seed-bank. Eventhough the addition of a seed-bank to the Fisher-Wright model is relatively new, in the past five years extensive research was done on extensions of the Fisher-Wright model with seed-bank. An overview of the state of the art is given in [54].

§1.2.3 The Fisher-Wright model with multi-layer seed-bank

A key idea in this thesis is that we can *enrich the seed-bank with internal states* to allow for fat tails and still preserve the Markov property for the evolution. To give the seed-bank an internal structure, we colour the dormant individuals with countably many colours $m \in \mathbb{N}_0$. Thus, instead of one seed-bank we have an infinite sequence of seed-banks, each with its own colour. Active individuals that become dormant are assigned a colour m at rate e_m. If an active individual is assigned a colour m, then it exchanges with a dormant individual of colour m. The colour m-dormant individual loses its colour when it becomes active, but retains its type. Therefore, during the evolution the relative size of the active population and the m-coloured seed-bank is fixed.

Evolution equation. Define

$$K_m = \frac{\text{size } m\text{-dormant population}}{\text{size active population}}, \qquad m \in \mathbb{N}_0, \qquad (1.19)$$

which denotes the relative size of the m-dormant population with respect to the active population. Let $(x(t))_{t \geq 0}$ denote the fraction of $\heartsuit$ in the active population at time t, and $(y_m(t))_{t \geq 0}$ the fraction of $\heartsuit$ in the m-dormant population at time t. So, we keep
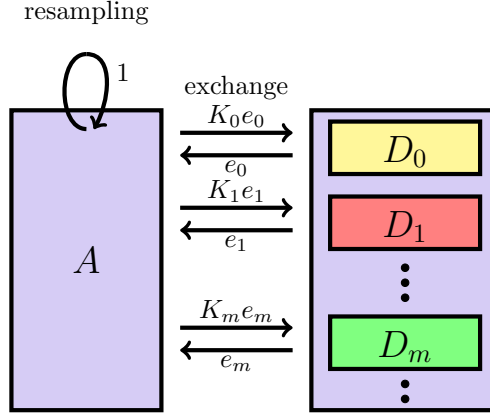
resampling

Figure 1.5: Schematic picture of the Fisher-Wright diffusion with layered seed-bank in (1.20). *Active individuals resample at rate 1, but exchange with a countable sequence of dormant populations. At rate $K_m e_m$ an active individual becomes dormant with colour m. An m-dormant individual becomes active at rate e_m.*

track of the complete sequence of dormant populations. In the continuum limit, the process $(x(t), (y_m(t))_{m \in \mathbb{N}_0})_{t \geq 0}$ evolves according to

$$\mathrm{d}x(t) = \sum_{m \in \mathbb{N}_0} K_m e_m \left[y_m(t) - x(t) \right] \mathrm{d}t + \sqrt{x(t)(1 - x(t))} \, \mathrm{d}w(t),$$

$$\mathrm{d}y_m(t) = e_m \left[x(t) - y_m(t) \right] \mathrm{d}t, \qquad m \in \mathbb{N}_0. \tag{1.20}$$

Comparing (1.20) to (1.13), we see that the active population exchanges with the whole sequence of dormant populations. However, each dormant population only evolves due to exchange with the active population. To ensure that active individuals do not become dormant instantaneously, we must assume that

$$\sum_{m \in \mathbb{N}_0} K_m e_m < \infty. \tag{1.21}$$

The evolution in (1.20) is depicted in Fig. 1.5.

Genealogy. Like for the Fisher-Wright model with seed-bank, we can describe the genealogy of the population. At a large time we sample from the population n active individuals, labeled by A, and k_m m-dormant individuals, labeled by D_m, for $m \in \mathbb{N}_0$. Then the lineages of the sampled individuals evolve according to the following rules.

- Each pair of active lineages coalesces at rate 1, independently of all other lineages.

- Each active lineage becomes m-dormant at rate $K_m e_m$.

- Each m-dormant lineage becomes active at rate e_m.

Similarly as for the Fisher-Wright model with (non-layered) seed-bank, we can define a layered seed-bank coalescent and a corresponding block-counting process.

Fat-tailed wake-up times. We define χ to be the total rate at which an active lineage becomes dormant, i.e.,

$$\chi = \sum_{m \in \mathbb{N}_0} K_m e_m. \tag{1.22}$$

Note that $\chi < \infty$ by (1.21). The distribution of the wake-up time τ defined in (1.17) for a lineage in the multi-layer seed-bank is given by

$$\mathbb{P}(\tau > t) = \sum_{m \in \mathbb{N}_0} \frac{K_m e_m}{\chi} \, \mathrm{e}^{-e_m t}. \tag{1.23}$$

Choosing the relative sizes of the seed-banks $(K_m)_{m \in \mathbb{N}_0}$ and the rates of exchange $(e_m)_{m \in \mathbb{N}_0}$ properly, we can mimic different wake-up time distributions. For example, we can choose

$$\begin{aligned} K_m &\sim A m^\alpha, \qquad e_m \sim B m^{-\beta}, \ m \to \infty \\ & A, B \in (0, \infty), \ \alpha, \beta \in \mathbb{R}, \ \alpha \le 1 < \alpha + \beta, \end{aligned} \tag{1.24}$$

where $\sim$ means asymptotically equal. Then

$$\mathbb{P}(\tau > t) \sim C t^{-\gamma}, \qquad t \to \infty, \tag{1.25}$$

where $\gamma = \frac{\alpha + \beta - 1}{\beta}$ and $C = \frac{A}{\chi \beta} B^{1-\gamma} \Gamma(\gamma)$, with Γ the Gamma-function. Therefore we can choose the sizes K_m and the rates e_m such that when we take the colours into account we still have a Markov process, but when we ignore the colours in the seed-bank we have a wake-up time that is fat-tailed.

Using the layered seed-bank model, we can choose the sequences $(K_m)_{m \in \mathbb{N}_0}$ and $(e_m)_{m \in \mathbb{N}_0}$ such that $0 < \gamma < \frac{1}{2}$ in (1.25). It turns out that in this case the Fisher-Wright diffusion with seed-bank in (1.20) no longer eventually reach its fixed points. (This was also observed in [11] for a non-Markovian seed-bank model.) Hence, in this parameter regime, *the layered seed-bank can prevent loss of genetic variability.*

A key quantity in the Fisher-Wright model with layered seed-bank is the relative size of the total seed-bank with respect to the active population:

$$\rho = \sum_{m \in \mathbb{N}_0} K_m. \tag{1.26}$$

The case $\rho = \infty$ shows different behaviour than the case $\rho < \infty$ (also this was observed in [11]). For example, for the expected wake-up time τ is we find

$$\mathbb{E}[\tau] = \frac{\rho}{\chi}. \tag{1.27}$$

In the next section we will turn to the layered Fisher-Wright model with seed-bank in the spatial setting. Also there the cases $\rho < \infty$ and $\rho = \infty$ give rise to qualitatively different long-term behaviour.

§1.3   Summary of Part I

In Part I of this thesis we consider a *spatial* version of the continuum Fisher-Wright models with seed-bank introduced in Sections 1.2.2–1.2.3. In the spatial version individuals live in colonies, each with their own seed-bank, and are allowed to *migrate* between colonies. The underlying geographic space is a countable Abelian group $\mathbb{G}$.

The spatial Fisher-Wright model without seed-bank has been the object of intense study. A sample of relevant papers and overviews is [66], [17], [20], [25], [22], [33], [29], [27], [24], [41]. In these papers the convergence of the system to equilibrium was proven. Parameter regimes were identified in which the spatial system converges to a mono-type equilibrium, i.e., the system grows locally mono-type clusters of only $\heartsuit$ or of only $\diamondsuit$, or in which the system converges to a multi-type equilibrium, i.e., locally both types are present. The first type of long-term behaviour is called *clustering*, the second type is called *coexistence*. It was shown that the dichotomy between clustering and coexistence for the spatial Fisher-Wright model without seed-bank is completely determined by the *migration kernel* according to which individuals migrate between colonies. If the migration kernel is transient, then coexistence prevails, while if the migration kernel is recurrent, then clustering prevails.

We expect that the presence of the seed-bank affects the long-time behaviour of the spatial system not only quantitatively but also qualitatively. To understand how this comes about, we must find ways to deal with the *richer behaviour* of the population caused by the motion in and out of the seed-bank. In [28] a spatial model with seed-bank, migration and mutation was analysed. There the probability for two individuals drawn randomly from two colonies to be identical by decent was computed as a function of the distance between the colonies.

The first goal in Part I is to prove convergence to equilibrium for the spatial Fisher-Wright model with seed-bank, and to identify the parameter regimes for clustering and coexistence. The second goal is to identify the role of the wake-up time. We will show that if the expected wake-up time is finite, then the dichotomy between clustering and coexistence is completely determined by the migration kernel and the seed-bank has only a quantitative effect on the long-term behaviour. However, if we allow the wake-up time to have infinite mean and moderately fat tails, then both the exchange rates with the seed-bank and the migration kernel determine the dichotomy. In that case the seed-bank has both a quantitative and a qualitative effect on the long-term behaviour. If the wake-up time has very fat tails, then the dichotomy is completely determined by the seed-bank, independently of the migration kernel.

In what follows we first introduce three models of increasing generality that are studied in Part I. After that we state the main results of Part I in words and briefly comment on the techniques used.

Basic ingredients for the models in Part I.   We extend the continuum Fisher-Wright models with seed-bank introduced in Sections 1.2.2–1.2.3 to three spatial models of increasing generality. In each of the three models, we consider populations of individuals of two types – either $\heartsuit$ or $\diamondsuit$ – located in colonies on a *geographic space* $\mathbb{G}$ that is a countable Abelian group. In each of the three models, the population in a

CHAPTER 1

colony consist of an *active* part and a *dormant* part. The *repository* of the dormant population at colony $i \in \mathbb{G}$ is called the seed-bank at $i \in \mathbb{G}$. Individuals in the active part of colony $i \in \mathbb{G}$ can resample, migrate and exchange with a dormant population. Individuals in the dormant part of colony $i \in \mathbb{G}$ can only exchange with the active population. An active individual that resamples chooses uniformly at random another individual from its colony and adopts its type. The rate of resampling can be state-dependent and is controlled by a diffusion function $g : [0,1] \to \mathbb{R}_{\geq 0}$ that satisfies the criteria for state-dependent resampling in Section 1.2.1. When an active individual at colony $i \in \mathbb{G}$ migrates, it chooses a parent from another colony $j \in \mathbb{G}$ and adopts its type. In each of the three models the migration is described by a *migration kernel* $a(\cdot, \cdot)$, which is an irreducible $\mathbb{G} \times \mathbb{G}$ matrix of transition rates satisfying

$$a(i,j) = a(0, j-i) \quad \forall \, i,j \in \mathbb{G}, \qquad \sum_{i \in \mathbb{G}} a(0,i) < \infty. \tag{1.28}$$

Here, $a(i,j)$ has to be interpreted as the rate at which an active individual at colony $i \in \mathbb{G}$ chooses an individual in the active part of colony $j \in \mathbb{G}$ and adopts its type. An active individual that becomes dormant *exchanges* with a randomly chosen dormant individual that becomes active. The dormant part of the population only evolves due to exchange of individuals with the active part of the population.

The three models we introduce below differ in the way the active population exchanges with the dormant population. However, in each of the three models the exchange mechanism guarantees that the sizes of the active and the dormant population stay fixed over time.

Since we look at continuum models obtained from individual-based models, we are interested in the fraction of individuals of type $\heartsuit$ in the different colonies.

Model 1: single-layer seed-bank. In this model we consider a multi-colony version of the continuum model in (1.13). Each colony $i \in \mathbb{G}$ has an active part A and a dormant part D. For $i \in \mathbb{G}$ and $t \geq 0$, let $x_i(t)$ denote the fraction of individuals in colony i of type $\heartsuit$ that are active at time t, and $y_i(t)$ the fraction of individuals in colony i of type $\heartsuit$ that are dormant at time t. Like in (1.10) in Section 1.2.2, let $K \in (0, \infty)$ be the relative size of the dormant population w.r.t. the active population, and let e be the rate at which active and dormant individuals exchange. We assume K and e to be the same for all colonies. The fractions of individuals of type $\heartsuit$ in the population evolve according to the system of stochastic differential equations (SSDE)

$$\mathrm{d}x_i(t) \;=\; \sum_{j \in \mathbb{G}} a(i,j) \left[x_j(t) - x_i(t) \right] \mathrm{d}t + \sqrt{g(x_i(t))} \, \mathrm{d}w_i(t) \tag{1.29}$$
$$+ \, Ke \left[y_i(t) - x_i(t) \right] \mathrm{d}t,$$

$$\mathrm{d}y_i(t) \;=\; e \left[x_i(t) - y_i(t) \right] \mathrm{d}t, \qquad i \in \mathbb{G}, \tag{1.30}$$

where $(w_i(t))_{t \geq 0}$, $i \in \mathbb{G}$, are independent standard Brownian motions. The first term in (1.29) describes the *migration* of active individuals from i to j at rate $a(i,j)$. The

second term in (1.29) describes the *resampling* of individuals at rate $\frac{g(x)}{x(1-x)}$ in state x. The third term in (1.29) together with the term in (1.30) describe the *exchange* of active and dormant individuals at rate $e \in (0, \infty)$.

Model 2: multi-layer seed-bank. In this model we consider a multi-colony version of the continuum model in (1.20). Therefore, for each $i \in \mathbb{G}$ a colony now consists of an active part A and a sequence $(D_m)_{m \in \mathbb{N}_0}$ of dormant parts, labeled by their colour $m \in \mathbb{N}_0$. As before, for $i \in \mathbb{G}$, let $x_i(t)$ denote the fraction of individuals in colony i of type $\heartsuit$ that are active at time t, but now let $y_{i,m}(t)$ denote the fraction of individuals in colony i of type $\heartsuit$ that are dormant with colour m at time t. Let $e_m \in (0, \infty)$ be the rate at which active individuals exchange with dormant individuals of colour m, and let $K_m \in (0, \infty)$ denote the relative size of the m-dormant population w.r.t. the active population as in (1.19). We assume e_m and K_m to be the same for all colonies. Then the fraction of $\heartsuit$ in the population evolves according to the SSDE

$$\mathrm{d}x_i(t) \;=\; \sum_{j \in \mathbb{G}} a(i,j)\left[x_j(t) - x_i(t)\right]\mathrm{d}t + \sqrt{g(x_i(t))}\,\mathrm{d}w_i(t) \tag{1.31}$$

$$+ \sum_{m \in \mathbb{N}_0} K_m e_m \left[y_{i,m}(t) - x_i(t)\right]\mathrm{d}t,$$

$$\mathrm{d}y_{i,m}(t) \;=\; e_m\left[x_i(t) - y_{i,m}(t)\right]\mathrm{d}t, \qquad m \in \mathbb{N}_0,\ i \in \mathbb{G}, \tag{1.32}$$

where $(w_i(t))_{t \geq 0}, i \in \mathbb{G}$, are independent standard Brownian motions. Comparing (1.31)–(1.32) with (1.29)–(1.30), we see that active individuals migrate (the first term in (1.31)) and resample (the second term in (1.31)) in the same way, but now interact with a sequence of dormant populations (the third term in (1.31) and the term (1.32)). The dormant individuals only exchange with the active individuals.

Model 3: multi-layer seed-bank with displaced seeds. We can extend the mechanism of Model 2 by allowing active individuals that become dormant to do so in a randomly chosen colony. This amounts to introducing a sequence of irreducible *displacement kernels* $a_m(\cdot, \cdot)$, $m \in \mathbb{N}_0$, satisfying

$$a_m(i,j) = a_m(0, j-i) \quad \forall\, i,j \in \mathbb{G}, \qquad \sum_{i \in \mathbb{G}} a_m(0,i) = 1 \quad \forall\, m \in \mathbb{N}_0, \tag{1.33}$$

and replacing (1.31)–(1.32) by

$$\mathrm{d}x_i(t) \;=\; \sum_{j \in \mathbb{G}} a(i,j)\left[x_j(t) - x_i(t)\right]\mathrm{d}t + \sqrt{g(x_i(t))}\,\mathrm{d}w_i(t) \tag{1.34}$$

$$+ \sum_{j \in \mathbb{G}} \sum_{m \in \mathbb{N}_0} K_m e_m\, a_m(j,i)\left[y_{j,m}(t) - x_i(t)\right]\mathrm{d}t,$$

$$\mathrm{d}y_{i,m}(t) \;=\; \sum_{j \in \mathbb{G}} e_m\, a_m(i,j)\left[x_j(t) - y_{i,m}(t)\right]\mathrm{d}t, \qquad m \in \mathbb{N}_0,\ i \in \mathbb{G}. \tag{1.35}$$

Here, the third term in (1.34) together with the term in (1.35) describe the *switch of colony* when individuals exchange between active and dormant. Namely, with

probability $a_m(i,j)$ simultaneously an active individual in colony i becomes dormant with colour m in colony j and a randomly chosen dormant individual with colour m in colony j becomes active in colony i.

Two key quantities. Like in Section 1.2.3, in Models 2 and 3 we must assume that

$$\chi = \sum_{m \in \mathbb{N}_0} K_m e_m < \infty \tag{1.36}$$

in order to make sure that active individuals do not become dormant instantly. Like in Section 1.2.3, in Models 2 and 3 we set

$$\rho = \sum_{m \in \mathbb{N}_0} K_m = \frac{\text{size dormant population}}{\text{size active population}}. \tag{1.37}$$

It turns out that χ and ρ are key quantities of the system. In particular, we will see that the long-time behaviour of Models 2 and 3 is different for $\rho < \infty$ and $\rho = \infty$.

Main results of Part I.

(1) For all three models the system convergences to a unique equilibrium that exhibits a dichotomy between *clustering* and *coexistence*. In all three models the density of $\heartsuit$ in the population is preserved over time, and in both the clustering case and the coexistence case the equilibrium depends on the initial density of $\heartsuit$. In the coexistence case the limiting equilibrium also depends on the resampling function, the exchange rates with the seed-bank and the migration kernel.

(2) For all three models we identify the parameter regimes for clustering and coexistence. These parameter regimes do not depend on the resampling function g.

   (2a) In Model 1 the wake-up time has *finite mean*. The dichotomy between coexistence and clustering is controlled by the migration only and the seed-bank has no effect on the dichotomy. In particular, clustering prevails when the symmetrised migration kernel is recurrent, while coexistence prevails when it is transient. This result is the classical dichotomy for populations without seed-bank [14].

   (2b) In Model 2 the wake-up time can have both *finite mean* and *infinite mean*. If the wake-up time has finite mean, then the dichotomy between coexistence and clustering is controlled by the migration only and the seed-bank has no effect, similarly as for Model 1. If the wake-up time has *infinite mean* with *moderately fat tails*, then the dichotomy is controlled by both the migration and the seed-bank. In particular, the parameter regimes for clustering and coexistence reveal an interesting interplay between rates for migration and rates for exchange with the seed-bank. If the wake-up time has *infinite mean* with *very fat tails*, then the dichotomy is controlled by the seed-bank only and the migration has no effect. For *infinite mean*

wake-up times it turns out that in the coexistence regime the seed-banks with colour $m \to \infty$ are in a state that is almost surely equal to the initial density of $\heartsuit$ in the population. Therefore *deep seed-banks become deterministic.*

(2c) In Model 3, the extra migration of active individuals that become dormant makes coexistence more likely. This extra migration can be incorporated into the dichotomy criterion obtained for model 2.

Techniques used in Part I. To prove the dichotomy, we first consider the three models when the resampling function is $g(x) = dx(1-x)$, $d \in \mathbb{R}_+$. For these diffusion functions we have duality relations similarly as those introduced in Section 1.2. As will be explained in detail in Part I, the lineages in the dual of the spatial model behave like a set of coalescing random walks. Afterwards, we can use comparison techniques to extend the results to a general resampling function g. To prove convergence to a unique equilibrium in the case of coexistence, we make use of *coupling* techniques.

§1.4 Summary of Part II

Part II of this thesis focuses on spatial populations with seed-bank where the underlying geographic space is the so-called *hierarchical group*, which we introduce next.

Hierachical lattice. The hierarchical lattice of order N is given by

$$\Omega_N = \left\{ \xi = (\xi_k)_{k \in \mathbb{N}_0} \colon \xi_k \in \{0, 1, \ldots, N-1\}, \ \sum_{k \in \mathbb{N}_0} \xi_k < \infty \right\}, \qquad (1.38)$$

which with addition modulo N becomes the hierarchical group of order N (see Fig. 1.6). The *hierarchical distance* is defined by

$$d_{\Omega_N}(\xi, \eta) = d_{\Omega_N}(0, \xi - \eta) = \min\left\{ k \in \mathbb{N}_0 \colon \xi_l = \eta_l \ \forall\, l \geq k \right\}, \qquad \xi, \eta \in \Omega_N. \quad (1.39)$$

Intuitively, depicting Ω_N as the leaves of an infinite tree as in Fig. 1.6, the distance between two points on Ω_N is the number of branches we have to travel upwards in the tree to find a common node.

The choice of Ω_N as geographic space plays an important role for population models, and was first exploited in [65] in an attempt to formalise ideas coming from ecology. One interpretation is that the sequence $(\xi_k)_{k \in \mathbb{N}_0}$ encodes the 'address' of colony ξ: ξ_0 is the 'house', ξ_1 is the 'street', ξ_2 is the 'village', ξ_3 is the 'province', ξ_4 is the 'country', and so on.

The goal of Part II of this thesis is to study the spatial Fisher-Wright model with seed-bank on the hierarchical lattice when the order N of the hierarchical group tends to infinity. This limit is called the *hierarchical mean-field limit*. To analyse the limiting system, it turns out that we have to consider different *space-time scales*. In what follows we first set up the model. After that we explain how the different space-time scales come into play in a natural way.
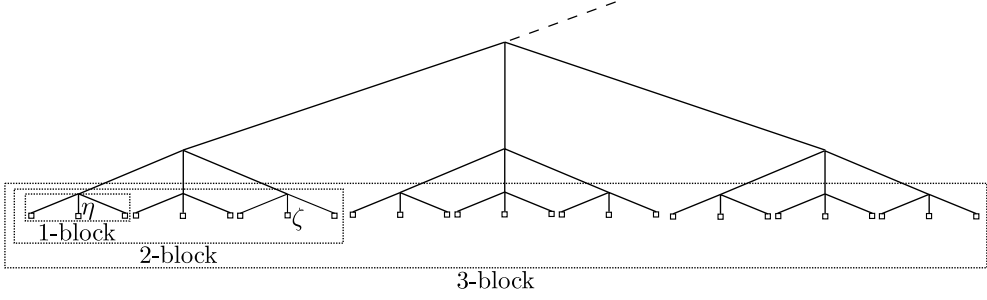
Figure 1.6: Close-ups of a 1-block, a 2-block and a 3-block in the hierarchical group of order $N = 3$. The elements of the group are the leaves of the tree (indicated by □'s). The hierarchical distance between two elements in the group is the graph distance to the most recent common ancestor in the tree: $d_{\Omega_3}(\eta, \zeta) = 2$ for η and ζ in the picture.

Hierarchical migration. We construct a migration kernel $a^{\Omega_N}(\cdot, \cdot)$ on the hierarchical group Ω_N built from a sequence of migration rates

$$\underline{c} = (c_k)_{k \in \mathbb{N}_0} \in (0, \infty)^{\mathbb{N}_0} \tag{1.40}$$

that do not depend on N. Individuals migrate as follows:

- For all $k \in \mathbb{N}$, each individual chooses at rate c_{k-1}/N^{k-1} the block of radius k around its present location and selects a colony uniformly at random from that block. Subsequently it selects an individual in this colony uniformly at random and adopts its type.

Since the block of radius k contains N^k colonies, the migration kernel is given by

$$a^{\Omega_N}(\eta, \xi) = \sum_{k \geq d_{\Omega_N}(\eta, \xi)} \frac{c_{k-1}}{N^{k-1}} \frac{1}{N^k}, \quad \eta, \xi \in \Omega_N, \eta \neq \xi, \qquad a^{\Omega_N}(\eta, \eta) = 0, \quad \eta \in \Omega_N.$$

$$\tag{1.41}$$

We assume that

$$\sum_{\xi \in \Omega_N} a^{\Omega_N}(\eta, \xi) < \infty \tag{1.42}$$

to guarantee that the total migration rate per individual is finite.

Evolution on the hierarchical lattice. The evolution of the single colonies in Part II is similar to the evolution of the single colonies in Part I in Model 2. The difference is that in the hierarchical setting we let the migration rates and the exchange rates depend on the order N of the hierarchical group. For $\xi \in \Omega_N$, define

$$x_\xi(t) = \text{ the fraction of active individuals of type } \heartsuit \text{ at colony } \xi \text{ at time } t,$$

$$y_{\xi,m}(t) = \text{ the fraction of } m\text{-dormant individuals of type } \heartsuit \text{ at colony } \xi \text{ and time } t.$$

$$\tag{1.43}$$

Active individuals exchange with dormant individuals with colour m at rate $\frac{e_m}{N^m}$, where $e_m \in (0, \infty)$. Like in (1.19), let K_m be the relative size of the m-dormant population with respect to the active population. Like in Part I, the sequences $(K_m)_{m \in \mathbb{N}_0}$

and $(e_m)_{m \in \mathbb{N}_0}$ are the same for all colonies. Also here we allow for a general diffusion function $g : [0,1] \to \mathbb{R}_{\geq 0}$ satisfying the conditions in Section 1.2.1.

The fraction of individuals of type $\heartsuit$ in the population evolves according to the SSDE

$$\mathrm{d}x_\xi(t) = \sum_{\eta \in \Omega_N} a^{\Omega_N}(\xi, \eta)[x_\eta(t) - x_\xi(t)] \, \mathrm{d}t + \sqrt{g(x_\xi(t))} \, \mathrm{d}w_\xi(t)$$

$$+ \sum_{m \in \mathbb{N}_0} \frac{K_m e_m}{N^m} [y_{\xi,m}(t) - x_\xi(t)] \, \mathrm{d}t, \tag{1.44}$$

$$\mathrm{d}y_{\xi,m}(t) = \frac{e_m}{N^m} [x_\xi(t) - y_{\xi,m}(t)] \, \mathrm{d}t, \quad m \in \mathbb{N}_0, \qquad \xi \in \Omega_N,$$

where we assume that $\sum_{m \in \mathbb{N}_0} \frac{K_m e_m}{N^m} < \infty$. The first term in the first equation describes the evolution of the active population at colony ξ due to migration, the second term due to resampling. The third term in the first equation and the term in the second equation describe the exchange between the active and the dormant population at colony ξ (see Fig. 1.7). Like the migration rates, the exchange rates between the active and the dormant population depend on the order N of the hierarchical group. Since dormant individuals are not subject to resampling or migration, the dynamics of the dormant population is completely determined by the exchange with the active population.
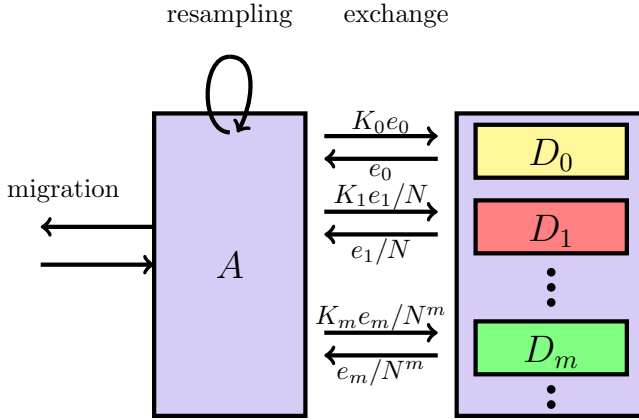

Figure 1.7: Active individuals (A) are subject to migration, resampling and exchange with dormant individuals (D). When active individuals become dormant they are assigned a colour $(D_m, m \in \mathbb{N}_0)$, which they lose when they become active again. The resampling rate in the active popluation depends on the type-$\heartsuit$ frequency x and equals $g(x)/x(1-x)$.

Evolution of block averages. The choice of the migration kernel in (1.41) implies that, for every $k \in \mathbb{N}$, at rate $\asymp \frac{1}{N^k}$ individuals choose a space horizon of distance $k+1$ and subsequently choose a random colony from that space horizon. Therefore, in order to see interactions over a distance $k+1$ for large N, we need to speed up time

by a factor N^k. A similar observation applies to the interaction with the seed-bank. Dormant individuals with colour k become active at rate $\asymp \frac{1}{N^k}$. Therefore, in order to see interactions with the k-dormant population for large N, we need to speed up time by a factor N^k. To analyse the effective interaction on different time scales N^k, $k \in \mathbb{N}_0$ we introduce *successive block averages*.

Definition 1.4.1. For $k \in \mathbb{N}_0$, let

$$B_k(0) = \{\eta \in \Omega_N \colon d_{\Omega_N}(0, \eta) \leq k\}$$

denote the k-block around 0. Define the k-block average around 0 at time $N^k t$ by

$$
\begin{aligned}
x_k^{\Omega_N}(t) &= \frac{1}{N^k} \sum_{\eta \in B_k(0)} x_\eta(N^k t), \\
y_{m,k}^{\Omega_N}(t) &= \frac{1}{N^k} \sum_{\eta \in B_k(0)} y_{\eta,m}(N^k t), \qquad m \in \mathbb{N}_0.
\end{aligned}
\tag{1.45}
$$

∎

The k-block average represents the dynamics of the system averaged over the N^k colonies around 0 with time speeded up by a factor N^k. Therefore we say that the k-block average represents the dynamics of the system *on space-time scale k*, or on *hierarchical level k*.

To obtain the *hierarchical mean-field limit*, we analyse the block averages defined above in the limit as $N \to \infty$, for which we expect a seperation of space-time scales. It turns out that in this limit each of the block averages preforms an autonomous diffusion, similar to the diffusion preformed by a single colony.

Main results Part II.

(1) For fixed N the results obtained in Part I are applied to the hierarchical model with seed-bank. For two classes of parameters the clustering regime is identified. In case the wake-up time has infinite mean, the clustering regime exhibits a trade off between the exchange rates and the migration rates.

(2) The hierarchical mean-field limit is identified, i.e., the evolution of the k-block averages defined in (1.45) is determined in the limit $N \to \infty$. For all $k \in \mathbb{N}_0$, the limiting k-block averages evolve according to a k-dependent SSDE, from which we can read off the following results:

(2a) For each $k \in \mathbb{N}_0$, if $N \to \infty$, then the migration terms in the evolution of the active k-block average can be replaced by a drift towards the active block one level up, i.e., a drift towards the active $(k + 1)$-block average. This phenomenon is called *decoupling*.

(2b) For each $k \in \mathbb{N}_0$, if $N \to \infty$, then the resampling rate of the active k-block average is the average resampling rate of the colonies within the k-block. The resampling rate of the active k-block is given by the k-fold

iteration of a *renormalisation transformation* $\mathcal{F}$ applied to the original diffusion function g. The resulting diffusion function $\mathcal{F}^{(k)}g$ is called the *renormalised diffusion function*.

(2c) For each $k \in \mathbb{N}_0$, if $N \to \infty$, then the active k-block exchanges only with the dormant k-block of colour k. Therefore we say that the k-dormant population is the *effective seed-bank* on level k. The k-block averages of dormant populations of colours $m < k$ start equalising with the active k-block averages and are called *fast seed-banks*. The seed-banks of colours $m > k$ do not change and stay in a fixed state. These seed-banks are refered to as *slow seed-banks*.

(2d) For each $k \in \mathbb{N}_0$, if $N \to \infty$, then the migration terms in the evolution of the active k-block averages, induce a drift towards the $(k+1)$-block average. It is in this way that subsequent block averages are connected. The connection between different hierarchical levels is captured by what is called the *interaction chain*.

(3) With the help of the interaction chain, the attracting orbit of the renormalisation transformation acting on the space of diffusion functions is analysed. In the *clustering regime* and after appropriate scaling, the renormalised diffusion function $\mathcal{F}^{(k)}g$ converges to the Fisher-Wright diffusion function as $k \to \infty$, irrespective of the diffusion function g controlling the resampling in the single colonies. This convergence shows that the hierarchical system exhibits *universality* on large space-time scales in terms of the scaling limit. For several subclasses of parameters the scaling of the renormalised diffusion function is identified. This scaling reveals a delicate interplay between the parameters controlling the migration and the seed-bank and also determines the speed at which *mono-type clusters grow in space and time*.

Techniques used in Part II. To prove the results in Part II we use the abstract schemes for mean-field analysis that were introduced in [21], [25] and [15]. To analyse the behaviour of the seed-bank, we use coupling results and a random walk interpretation of the system that is proven in Part I. For the analysis of the renormalisation transformation we proceed like in [5].

§1.5   Further research

Two topics that would be interesting to study are the finite-systems scheme and the genealogy and cluster formation.

Finite-systems scheme. In Part I we analyse a spatial version of the Fisher-Wright model with seed-bank where the underlying space is a countable Abelian group. A key question is how well the infinite systems introduced in Part I can approximate "real-world" finite systems. To answer this question, we use the so-called finite-systems scheme that was introduced in [15]. In the *finite-systems scheme* we trucate both the

geographic space and the seed-bank, and subsequently let both the truncation level and the time tend to infinity, properly tuned together.

In an upcoming paper [44], we focus on Model 2, introduced in Section 1.3 and evolving according to (1.31)–(1.32). We study the truncated system in the parameter regime of coexistence for the non-truncated system. This parameter regime is identified in one of the main theorems in Part I. For truncated finite systems, we always get clustering. To obtain a meaningful scaling limit, the truncation level and the time are scaled in such a way that we are just at the verge of seeing the clustering of the finite system coming in. We find that, with this scaling, the finite system behaves like the infinite system in equilibrium in the coexistence regime, but with a density of $\heartsuit$ that is random.

Again we see a difference in behaviour between $\rho < \infty$ and $\rho = \infty$. In case $\rho < \infty$, we can adapt classical techniques to analyse the finite-systems scheme and the results lead to a single universality class. In case $\rho = \infty$, we can tune the speed at which the seed-bank tends to infinity relative to the speed at which the geographic space tends to infinity. This leads to new phenomena, and different universality classes appear.

Cluster formation and genealogy. It would be interesting to study the growth of monotone clusters in the spatial setting. A short introduction to cluster formation will be given in the setting of the hierarchical group in Part II. However, it would be interesting to see what the effect of the seed-bank is on the cluster formation. Closely related to the cluster formation is the genealogy. Scaling of the genealogy leads to continuum random trees with dormancy. We have not yet started this part of the research.

§1.6   Outline of the thesis

Part I of this thesis is based on [43]. It treats the general setting of spatial populations with seed-bank. In Chapter 2 we formally introduce the models of Section 1.3, and subsequently state the main results about well-posedness, duality and long-term behaviour. Chapter 3 is devoted to the proofs of the main results stated in Chapter 2.

Part II is based on the upcoming paper [45]. It treats the spatial hierarchical seed-bank model introduced in Section 1.4. In Chapter 4 we formally introduce the hierarchical seed-bank model and subsequently state the main results about multi-scaling renormalisation and universality. Chapters 5–10 are devoted to the proofs of the main results stated in Chapter 4.