Modeling the evolution of spatially distributed populations

Raphaël Forien

September 25, 2014

Introduction and historical sketch

In 1859, Charles Darwin issued his landmark book *On The Origin of Species*, expounding his views and evidence on the evolution of species by means of natural selection. Fixism - the theory that species do not evolve - had already been strongly criticised earlier in the 19^{th} century, but Darwin, along with Alfred Russel Wallace, was the first to describe how the struggle for survival led to natural selection and the accumulation of small profitable variations, thus accounting for the evolution of species.

Much debate ensued in the following decades and even though the scientific community became rapidly convinced that evolution was indeed taking place, many differed on its chief mechanism. Apart from the advocates of natural selection, some favored what would later be known as neo-Lamarckism, after the French naturalist and early defender of the transmutation of species. Neo-Lamarckism was the idea that inheritance of acquired characteristics, developed through use and disuse (as muscles are) produced the observed adaptations in most species. Others held that evolution took place through radical mutation events resulting in the forming of new species. This view, called saltationism, was popular among the first geneticists.

This debate reflected the lack of knowledge on the laws of heredity and the mechanisms producing individual variations at the time. Incidentally, these laws were uncovered in a series of experiments by the now famous Gregor Mendel at around the same time that Darwin published the first version of the *Origin of Species*, but remained forgotten until 1900. The rediscovery of Mendel's work resulted in a chasm among biologists of the early 20^{th} century. Inheritance of acquired characteristics, along with Lamarckism, fell out of favor and the community was divided between Mendelians, studying evolution by means of the inheritance of discrete mutations, and the others who argued in favor of gradual variation of heritable traits which, in their view, could not be accounted for by Mendelian genetics. Further evidence strongly supporting Mendel's laws was provided by Morgan's early 1910s experiments on the fruit fly, during which he succeeded to observe Mendelian mechanisms in the inheritance of various mutations, but failed to produce any mutation large enough to give rise to a new species on its own.

The controversy lasted until the 1920s and the publishing of the works of Ronald A. Fisher, John B. S. Haldane and Sewall Wright, setting the cornerstone of population genetics. They developed new mathematical models based on Mendelian inheritance and showed how natural selection can act on gene frequencies, resulting in small, progressive, variation as mutations become fixed and accumulate in a given population. This reconciliation of Mendelian genetics and natural selection came to be known as the modern evolutionary synthesis and established the legitimacy of evolutionary biology throughout the 1920s and 1930s.

Later developments, with the rise of molecular biology in the 1950s, the discovery of DNA and of its function as a code for the manufacturing of proteins, helped to better understand the mechanisms of inheritance and evolution. In the 1960s, under the influence of W. D. Hamilton, John Manyard Smith, the gene became the conceptual protagonist of natural selection - whereas, up to this time, arguments involving "the good of the species" or group selection were still widely used in evolutionary biology. The gene-centered view of natural selection can be summed up as follows : a gene or a mutation coding for a change of behaviour which favors the increasing of its own frequency within a given population tends to spread in this population. This is most commonly achieved by increasing the mean number of offspring of the gene's bearer, in the primary meaning of natural selection, but it can also be otherwise, for example by favoring the survival and reproduction of related individuals - who share the same gene with some probability. This particular kind of natural selection, called kin selection, accounts mainly for altruism and social behaviour among relatives of a given species. From this point, the conceptual core of the theory became relatively stable, with the exception of the discovery of horizontal gene transfer and epigenetic inheritance.

We are here interested in presenting some particular models arising in population genetics, aiming to describe populations evolving in a spatial continuum. These models reflect the highly probabilistic flavour of evolution :

random reproductive success of individuals results in genetic drift and random evolution of gene frequencies. We will not include natural selection in the models presented here, but it is worth keeping in mind that it can be taken into account easily (see [Eth09]). These models are overly simplistic regarding many aspects, but they give a very general description of the mechanisms of evolution at the level of gene frequencies in a given population.

We will start with one of the simplest models in population genetics : the Moran model, and outline some of its properties. Then we shall turn to a process describing the genealogy of a subsample of the modeled population as one looks back in time. Finally, we will present a recently introduced framework for modeling populations evolving in a spatial continuum.

1 The Moran model

Any proper introduction to the subject of population genetics ought to start by describing the Wright-Fisher model, and then proceed to more sophisticated models. However, for the sake of brevity, we chose here to omit the Wright-Fisher model and instead present its continuous time counterpart, the Moran model. The latter is indeed easier to handle from the mathematician's point of view while displaying similar behaviours. A (proper) introduction to all these models and their properties related to population genetics can be found in [Eth09].

Definition of the Moran model Let us consider a population of N haploid individuals (*i.e.* each individual carries a single copy of each gene, unlike diploid organisms such as us) which reproduce by simply producing copies of themselves (no sexual reproduction). We make several fundamental assumptions, namely that the population is panmictic (*i.e.* there is no internal structure within the population, in particular the effects of its spatial distribution are negligible) and its size is constant. Each individual in the population carries one or the other of two possible alleles of one particular gene. We denote the two alleles 0 and 1 and X_t will stand for the frequency of type 1 individuals in the population at time t (*i.e.* if there are k individuals in the population carrying allele 1 at time t, then $X_t = \frac{k}{N}$).

Given the initial state of the population, let $(\Pi_t)_{t\geq 0}$ be a Poisson process with intensity $\binom{N}{2}$. At each jump time of $(\Pi_t)_{t\geq 0}$, a pair of individuals is sampled (uniformly at random) in the current population, one of them dies and the other produces a copy of itself (it reproduces), each with equal probabilities.

We are interested in the behaviour of $(X_t)_{t\geq 0}$. From the definition above, it inherits the Markov property from $(\Pi_t)_{t\geq 0}$. At each jump time (which occur according to the Poisson process $(\Pi_t)_{t\geq 0}$), it can either increase by $\frac{1}{N}$ or decrease by $\frac{1}{N}$, each with probability $X_{t^-}(1-X_{t^-})$ - where t is the jump time - else it remains constant. Thus it is also a (continuous time) martingale (for which most of the results on discrete time martingales hold). In addition, its state space is finite and has two absorbing states : 0 and 1. As a consequence, $(X_t)_{t\geq 0}$ will eventually reach one of these in finite time and we then say that the corresponding allele has reached fixation in the population. Doob's optional sampling theorem tells us that the probability of reaching 1 before 0 is exactly X_0.

Infinitesimal generator We now need to introduce briefly a very useful tool in the study of Markov processes. Let $(X_t)_{t\geq 0}$ be a Markov process with state space E. The infinitesimal generator A of $(X_t)_{t\geq 0}$ is defined as follows. For any function $f : E \to \mathbb{R}$ and $x \in E$, let

$$Af(x) = \lim_{\delta t \downarrow 0} \frac{1}{\delta t}\mathbb{E}\left[f(X_{\delta t}) - f(X_0) \mid X_0 = x\right]$$

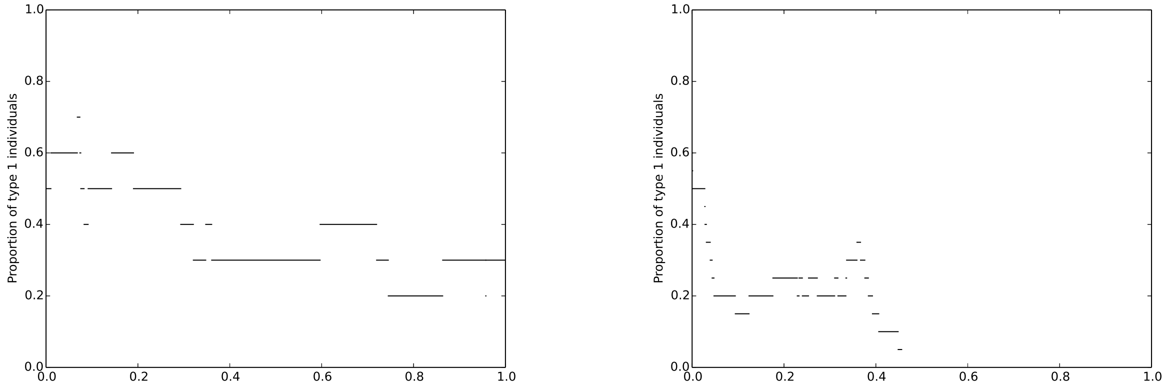
whenever the limit exists. The set of functions such that the limit exists is called the domain of A. It can be seen as a probabilistic derivative of the process $(X_t)_{t\geq 0}$, and indeed it retains information about the instantaneous behaviour of the process.

Let us look at the example of a Poisson process with intensity λ. In that case, conditioning on the value of $X_{\delta t}$ and provided f is bounded, we can write :

$$\begin{aligned}\mathbb{E}_x\left[f(X_{\delta t})\right] &= f(x)e^{-\lambda\delta t} + f(x+1)e^{-\lambda\delta t}\lambda\delta t + o\left(\delta t\right) \\ &= f(x)(1-\lambda\delta t) + f(x+1)\lambda\delta t + o\left(\delta t\right).\end{aligned}$$

From which we can deduce that

$$Af(x) = \lambda\left(f(x+1) - f(x)\right).$$



(a) (b)

Figure 1: Sample paths of $(X_t)_{t\geq 0}$.
Here is what the trajectory of $(X_t)_{t\geq 0}$ can look like for (a) $N = 10$ and $X_0 = 0.5$, (b) $N = 20$ and $X_0 = 3/5$.

This form for the infinitesimal generator is typical of jump processes. Poisson processes make only jumps of size one at a constant rate (here denoted λ) but a variety of situations could be handled in a similar way. For example one can show as an exercise that the infinitesimal generator of the process $(X_t)_{t\geq 0}$ defined in the Moran model is the following :

$$Af(x) = \frac{N^2}{2}x(1-x)\left(f(x+\frac{1}{N}) - f(x)\right) + \frac{N^2}{2}x(1-x)\left(f(x-\frac{1}{N}) - f(x)\right). \tag{1.1}$$

Another important example is that of Brownian motion. Taking f to be twice continuously differentiable on $\mathbb{R}$ and using Taylor's expansion theorem, one can show that the generator of standard Brownian motion takes the form :

$$Af(x) = \frac{1}{2}f''(x). \tag{1.2}$$

Infinitesimal generators come in handy when one wants to characterize the law of a process. Indeed, provided the domain of the generator of a process $(X_t)_{t\geq 0}$ is large enough, any Markov process having the same generator is a version of $(X_t)_{t\geq 0}$. See [EK09, Chapter 4.4] for a proof of this result.

One widespread application of this is to characterize the possible limits of a sequence of Markov processes. More precisely, let $\left\{\left(X_t^N\right)_{t\geq 0}, N \geq 1\right\}$ be a sequence of Markov processes and denote by A^N the generator of $\left(X_t^N\right)_{t\geq 0}$. Also take $(X_t)_{t\geq 0}$ to be a Markov process with generator A. Suppose that, for a sufficiently large set of functions f, and for all $x \in E$, $A^N f(x)$ converges to $Af(x)$. Then, under suitable conditions (see [EK09, Chapter 4.8]), for all $f : E \to \mathbb{R}$ and all $0 \leq t_1 < \ldots < t_k$,

$$\left(f(X_{t_1}^N), \ldots, f(X_{t_k}^N)\right) \underset{N\to\infty}{\longrightarrow} \left(f(X_{t_1}), \ldots, f(X_{t_k})\right)$$

in distribution ; we then say that the finite dimensional distributions of $\left(X_t^N\right)_{t\geq 0}$ converge to that of $(X_t)_{t\geq 0}$. Note that this is not the same as the convergence of the distribution of the whole process. In the case of *càdlàg* processes (*i.e.* whose sample paths are almost surely right-continuous with left limits), it is often possible to show that the sequence of probability measures on the space of *càdlàg* applications corresponding to the distributions of $\left(X_t^N\right)_{t\geq 0}$ is relatively compact, using results detailed in [EK09, Chapter 3]. If this is the case, then the convergence of the finite dimensional distributions is enough to ensure the convergence of the distribution of the whole process $\left(X_t^N\right)_{t\geq 0}$, since there is at most one *càdlàg* process with the same finite dimensional distributions as $(X_t)_{t\geq 0}$.

Large population size Let us now use this approach to describe the behaviour of the Moran model when the population size is assumed to be infinitely large. For $N \geq 1$, let $\left(X_t^N\right)_{t\geq 0}$ be defined as above to be the frequency of type 1 individuals in a population of N individuals evolving according to the Moran model. We take f to be twice continuously differentiable. Writing a Taylor expansion for f in (1.1), we obtain, for the infinitesimal generator of $\left(X_t^N\right)_{t\geq 0}$:

$$A^N f(x) = \frac{1}{2}x(1-x)f''(x) + \mathcal{O}\left(\frac{1}{N}\right).$$

Thus the sequence of generators admits a limit, and the finite dimensional distributions of $\left(X_t^N\right)_{t\geq 0}$ converge to that of the process whose generator is

$$Af(x) = \frac{1}{2}x(1-x)f''(x). \tag{1.3}$$

Note the similarity with the generator of standard Brownian motion (1.2). Indeed these two processes belong to a wider class of random processes called one-dimensional diffusions. In fact we have the following stronger result.

Theorem 1.1 (Diffusion approximation for the Moran model)**.** *There exists a process* $\left(X_t\right)_{t\geq 0}$ *such that*

$$\left(X_t^N\right)_{t\geq 0} \underset{N\to\infty}{\longrightarrow} \left(X_t\right)_{t\geq 0}$$

in distribution and whose infinitesimal generator is (1.3).

Without entering too much into detail, we can describe quite easily this limiting process, called the Wright-Fisher diffusion (it first arose as a limiting process in the Wright-Fisher model).

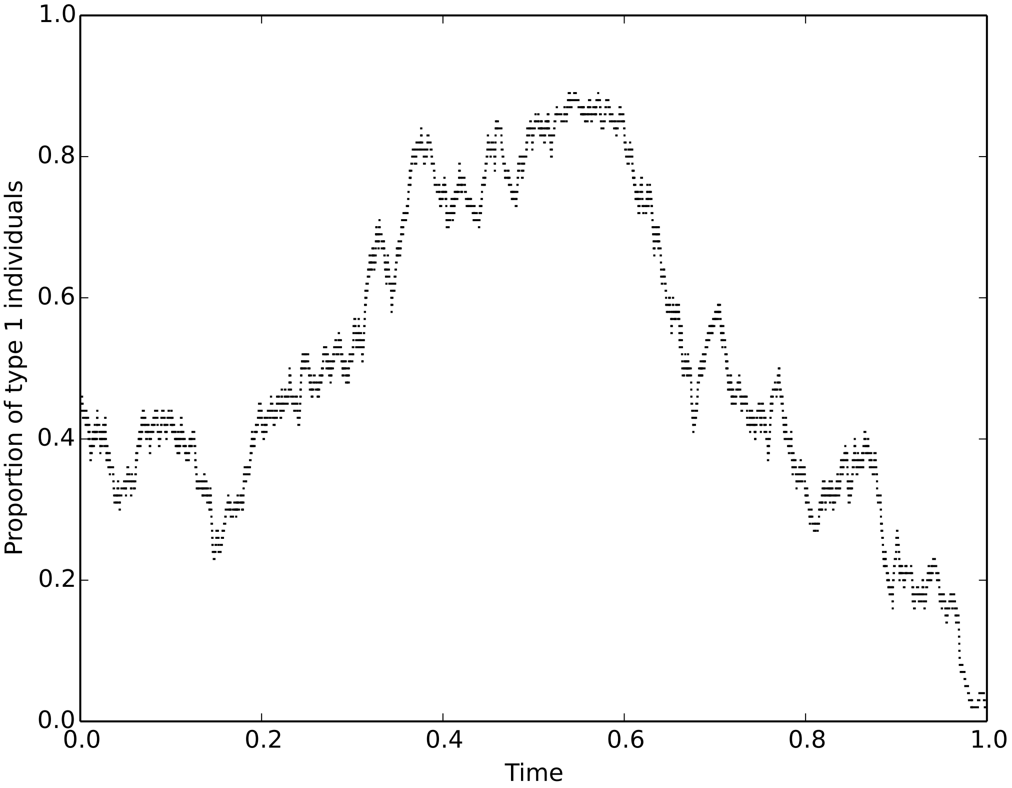


Figure 2: Trajectory for large population size.
Here is a sample path for $N = 100$, giving a good idea of the shape of the limiting diffusion.

First consider the case $f(y) = y$, the definition of the infinitesimal generator yields :

$$\mathbb{E}_x \left[X_{\delta t} - x \right] = o\left(\delta t\right).$$

This implies that the limiting process is also a martingale. Note that for general diffusion processes, this term is not necessarily zero - as is the case when one takes into account mutations and selection ; it is called the infinitesimal drift of the process. Now taking $f(y) = (y - x)^2$ yields :

$$\mathbb{E}_x \left[(X_{\delta t} - x)^2 \right] = x(1 - x)\delta t + o\left(\delta t\right).$$

This is why we call $x(1-x)$ the infinitesimal variance of the diffusion $(X_t)_{t \geq 0}$. Note that because it is zero for either $x = 0$ or $x = 1$, the diffusion will stop upon hitting the boundary $\{0, 1\}$. This variance measures what biologists call genetic drift, not to be mistaken for the infinitesimal drift of the process which instead measures the effect of mutations and natural selection. Figure 2 suggests - and it can be made rigorous, in fact it is the case of all diffusion processes - that the path $t \mapsto X_t$ is continuous almost surely. The following result shows that $(X_t)_{t \geq 0}$ can be seen as a (randomly) time changed Brownian motion.

Theorem 1.2. *For $x \in (0, 1)$, let $(B^x_t)_{t \geq 0}$ be a standard Brownian motion started at x. Define $T^{0,1} = \inf\{s \geq 0 : B^x_s \in \{0, 1\}\}$ and, for $t < T^{0,1}$,*

$$\tau(t) = \int_0^t \frac{\mathrm{d}s}{B^x_s (1 - B^x_s)}.$$

Since τ is monotonous and continuous, it is a bijection from $[0, T^{0,1})$ to $[0, c)$ for some $0 \leq c \leq \infty$. Thus we can define $m(t) = \tau^{-1}(t)$ for $t \in [0, c)$. Then the process $(X_t)_{t \geq 0}$ defined by $X_t = B^x_{m(t)}$ has the following infinitesimal generator :

$$Af(x) = \frac{1}{2} x(1 - x) f''(x).$$

Proof. First note that

$$m'(t) = \frac{1}{\tau'(m(t))} = X_t (1 - X_t).$$

Recall that $(B^x_t - B^x_0)^2 - t$ defines a martingale with respect to the natural filtration $(\mathcal{F}_t)_{t \geq 0}$ of $(B^x_t)_{t \geq 0}$. Checking that, for all $s \in [0, c)$, $m(s)$ is a $(\mathcal{F}_t)_{t \geq 0}$ stopping time, we can use Doob's optional sampling theorem to write, for $\delta t > 0$:

$$\begin{aligned}
\mathbb{E}\left[\left(B^x_{m(\delta t)} - B^x_0 \right)^2 - m(\delta t) \right] &= 0 \\
\mathbb{E}\left[\left(B^x_{m(\delta t)} - B^x_0 \right)^2 \right] &= m'(0)\delta t + o\left(\delta t\right) \\
&= x(1 - x)\delta t + o\left(\delta t\right).
\end{aligned}$$

Likewise one can show that $\mathbb{E}\left[B^x_{m(\delta t)} \right] = x$. Then, using Taylor's theorem, we can write :

$$\mathbb{E}_x \left[f(X_{\delta t}) - f(X_0) \right] = \frac{1}{2} f''(x) \mathbb{E}\left[\left(B^x_{m(\delta t)} - B^x_0 \right)^2 \right] + o\left(\delta t\right).$$

Dividing by δt and letting $\delta t \downarrow 0$ yields the result. □

The study of the process $(X_t)_{t \geq 0}$ provides information on the evolution of the frequency of a given allele forwards in time, but what of the genealogy of a sample of individuals taken at random within the population ? The answer to this is given in the next section.

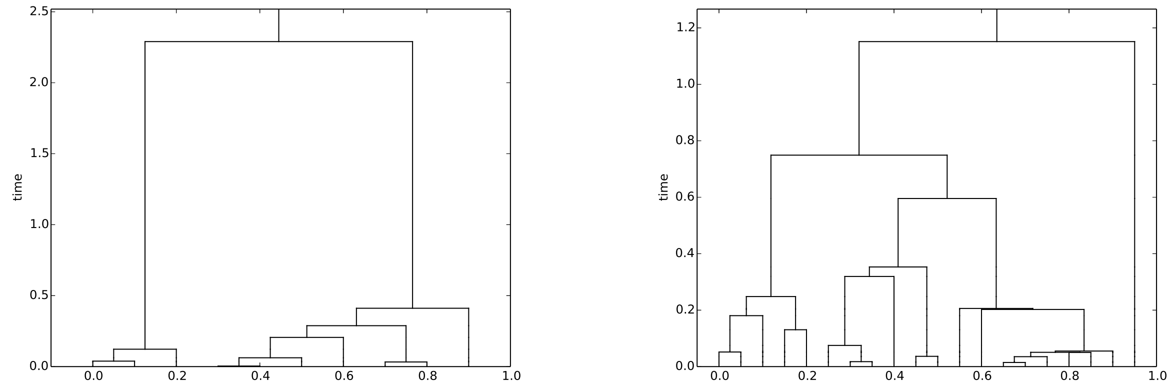
2 Kingman's coalescent

Picture yourself a sample of - say k - drops of water moving haphazardly on some water-repellent surface. After some time two drops will merge - or coalesce - to form one (bigger) drop of water, and so on, until there remains only one big drop (but is it still a *drop* then ?). Similarly, consider k individuals in a population evolving according to the Moran model at time t, and retrace their past ancestry. At some time before t, two of these individuals will have the same ancestor (because they were born in the same reproduction event). Looking even further in the past, the set of the ancestors of our sample will eventually be reduced to one common ancestor. The common feature of these two situations is what is captured by the notion of coalescent processes.

In the following, if Π is a partition of $\{1, \dots, N\}$, we will denote by $\{A_1, \dots, A_k\}$ the blocks of the partition. In case it is necessary, recall that they are disjoint non-empty subsets of $\{1, \dots, N\}$ whose union is the whole of $\{1, \dots, N\}$. If $m \leq N$, Π naturally induces a partition of $\{1, \dots, m\}$ whose blocks are the restrictions of the A_i to $\{1, \dots, m\}$, we call this partition the trace of Π on $\{1, \dots, m\}$.

Definition 2.1 (Kingman's coalescent)**.** *Let* $N \geq 1$ *and* $(\Pi_t)_{t\geq 0}$ *be a Markov process whose state space is the set of partitions of* $\{1, \dots, N\}$. *Suppose that* $(\Pi_t)_{t\geq 0}$ *is such that, given its initial state* $\{A_1, \dots, A_k\}$, *let* $(\tau_{i,j})_{1\leq i<j\leq k}$ *be independent random variables, exponentially distributed with parameter 1 ; then if* $\tau_{i,j} = \min_{k,l} \tau_{k,l}$, *at time* $t = \tau_{i,j}$, Π_t *becomes the partition whose blocks are obtained by taking the blocks of* Π_0 *and merging* A_i *and* A_j, *leaving the others untouched. The behaviour of the process after this first jump is determined by proceeding recursively, making use of the Markov property.* $(\Pi_t)_{t\geq 0}$, *when starting from the trivial partition in* N *singleton, is called Kingman's coalescent.*

It is easily shown that there is always a unique pair (i, j) such that $\tau_{i,j} = \min_{k,l} \tau_{k,l}$, and thus an equivalent description of Kingman's coalescent is to say that after an exponentially distributed time with parameter $\binom{k}{2}$ where k is the current number of blocks in Π_t, two blocks sampled at random in Π_t merge.



(a) (b)

Figure 3: Examples of Kingman coalescents.
Coalescents can be represented using trees where, each time two blocks happen to merge, their corresponding branches are linked to a node. Here are two such pictures with $N = 10$ for (a) and $N = 20$ for (b) (time flows upwards).

A number of things can be said of coalescent processes and in particular Kingman's coalescent, but our aim is not to introduce in full this topic. One can find more about coalescents in [Pit06]. Let us simply mention a few facts that are relevant from the biological point of view. First, the time we have to wait before two given elements become members of the same block is exponentially distributed with parameter 1 (careful, the pictures given in Figure 2 can be misleading as we have chosen a particular embedding of the tree *after* its structure had been decided). Also, if $m \leq N$, the trace of Π_t on $\{1, \dots, m\}$ is again a Kingman's coalescent on $\{1, \dots, m\}$. Kolmogorov's extension theorem then ensures that there exists a process $(\Pi_N^\infty)_{N\geq 1}$ such that, for each $N \geq 1$, the law of $(\Pi_N^\infty(t))_{t\geq 0}$ is that of Kingman's coalescent on $\{1, \dots, N\}$ and, if $m \leq N$, Π_m^∞ is the trace of Π_N^∞ on $\{1, \dots, m\}$. $(\Pi^\infty(t))_{t\geq 0}$ is then a Markov process whose state space is the space of partitions of $\mathbb{N}$, this process is also called Kingman's coalescent.

Moment duality with the Wright-Fisher diffusion There is a very deep connection between Kingman's coalescent and the Wright-Fisher diffusion described previously. We shall outline one aspect of this connection by the study of moment duality.

Definition 2.2 (Duality)**.** *Let* $(X_t)_{t\geq 0}$ *be a stochastic process defined on a probability space* $\left(\Omega, \mathcal{F}, \overrightarrow{\mathbb{P}}\right)$ *taking values in some space* E_1 *and let* $(Y_t)_{t\geq 0}$ *be another stochastic process defined on a space* $\left(\Omega', \mathcal{F}', \overleftarrow{\mathbb{P}}\right)$ *and taking values in* E_2*. Let* $h : E_1 \times E_2 \to \mathbb{R}$ *be a measurable function such that :*

$$\forall t \geq 0,\ \forall (x,y) \in E_1 \times E_2,\ \ \overrightarrow{\mathbb{E}}_x\left[h(X_t, y)\right] = \overleftarrow{\mathbb{E}}_y\left[h(x, Y_t)\right]. \tag{2.1}$$

Then $(X_t)_{t\geq 0}$ *and* $(Y_t)_{t\geq 0}$ *are said to be dual with respect to* h*.*

One should think of $(X_t)_{t\geq 0}$ as a forwards-in-time process and $(Y_t)_{t\geq 0}$ as a backwards-in-time process, as indicated by the little arrows on top of the corresponding expectancies. A more general definition of duality, along with more detailed results, is given in [EK09, Chapter 4.4, Theorem 11]. We have the following characterisation in terms of infinitesimal generators.

Proposition 2.3. *Let* $\overrightarrow{A}$ *and* $\overleftarrow{A}$ *denote the infinitesimal generators of* $(X_t)_{t\geq 0}$ *and* $(Y_t)_{t\geq 0}$*, respectively. Then* $(X_t)_{t\geq 0}$ *and* $(Y_t)_{t\geq 0}$ *are dual with respect to* h *if and only if*

$$\forall (x,y) \in E_1 \times E_2,\ \ \overrightarrow{A}h(x,y) = \overleftarrow{A}h(x,y).$$

Sketch of proof. Assuming that (2.1) holds, one writes :

$$\begin{aligned}
\overrightarrow{A}h(x,y) &= \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \overrightarrow{\mathbb{E}}_x\left[h(X_{\delta t}, y) - h(x,y)\right] \\
&= \lim_{\delta t \downarrow 0} \frac{1}{\delta t} \overleftarrow{\mathbb{E}}_y\left[h(x, Y_{\delta t}) - h(x,y)\right] \\
&= \overleftarrow{A}h(x,y),
\end{aligned}$$

yielding the first part. To prove (2.1), the idea is to show that :

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\left[h(X_s, Y_{t-s})\right] &= \mathbb{E}\left[\overrightarrow{A}h(X_s, Y_{t-s}) - \overleftarrow{A}h(X_s, Y_{t-s})\right] \\
&= 0.
\end{aligned}$$

Integrating between 0 and t yields the result. □

After seeing our notations, one will have guessed what we are coming to : the Wright-Fisher diffusion and Kingman's coalescent are dual with respect to some function.

Theorem 2.4 (Moment duality with the Wright-Fisher diffusion)**.** *Let* $(X_t)_{t\geq 0}$ *be the stochastic process defined in the neutral Moran model with infinite population size of section 1 and let* $(\Pi_t)_{t\geq 0}$ *be Kingman's coalescent. We denote by* N_t *the number of blocks in* Π_t *(we haven't specified the initial number of those blocks, but that will be done later). For all* $n \geq 1$ *and* $x \in [0,1]$*, we have the following duality relation :*

$$\overrightarrow{\mathbb{E}}_x\left[X_t^n\right] = \overleftarrow{\mathbb{E}}_n\left[x^{N_t}\right], \tag{2.2}$$

where $\overrightarrow{\mathbb{E}}$ *is the expectancy with respect to* $(X_t)_{t\geq 0}$ *and* $\overleftarrow{\mathbb{E}}$ *is with respect to* $(\Pi_t)_{t\geq 0}$ *(*$\overleftarrow{\mathbb{E}}_n$ *means that* $(\Pi_t)_{t\geq 0}$ *starts with* n *blocks).*

Proof of Theorem 2.4. The function we consider is $h(x,n) = x^n$. The infinitesimal generator of the Wright-Fisher diffusion has been derived in (1.3). Hence, for $x \in [0,1]$ and $n \geq 1$,

$$\begin{aligned}
\overrightarrow{A}h(x,n) &= \frac{1}{2}x(1-x)\frac{\partial^2 h}{\partial x^2}(x,n) \\
&= \frac{1}{2}(1-x)n(n-1)x^{n-1} \\
&= \binom{n}{2}\left(h(x, n-1) - h(x,n)\right) \\
&= \overleftarrow{A}h(x,n).
\end{aligned}$$

Theorem 2.4 then follows from Proposition 2.3, since $(N_t)_{t\geq 0}$ is a Markov process which jumps from n to $n-1$ after an exponential time with parameter $\binom{n}{2}$. $\square$

The left hand side of (2.2) is the probability that n individuals sampled in the infinite population corresponding to the Wright-Fisher diffusion at time t are all of type 1. On the right hand side of (2.2), we have the probability that all the ancestors of a sample of size n whose genealogy is given by Kingman's coalescent were of type 1 at time t before present. When we say that the genealogy of a sample is described by Kingman's coalescent, we mean that the blocks of Π_t are the subsets of individuals of the sample sharing a common ancestor at time t before present. Thus Theorem 2.4 simply says that - in a weak sense, but finer results in fact hold - the genealogy of a sample taken in the Moran model with infinite population size evolves (backwards) according to Kingman's coalescent. It is shown in [DK96] that both processes can be constructed on the same probability space, using a so-called lookdown construction, providing the aforementioned stronger duality.

3 Introducting spatial structure

The Moran model has some nice mathematical properties, but it is far too simplistic and by no means provides for a satisfactory description of any real population. One key feature that this model lacks is spatial structure. Most populations are scattered across some geographical space and thus fail to satisfy our assumption that the population is panmictic. Indeed there is a long history of attempts at modeling spatially structured populations. The following is based on the account given in [BEV13].

One of the earliest models taking spatial structure into account is Wright's island model [Wri43]. Wright considers a population equally divided into islands - or demes - with migration occurring between demes. More precisely, the subpopulation within each island evolves according to the Wright-Fisher model, at each generation, individuals in all the demes reproduce and a fixed proportion m of the offspring migrates to a randomly chosen island.

Motoo Kimura, a renowned population geneticist, further developed this setting by considering more general graphs (*e.g.* $\mathbb{Z}^2$). For all pairs (i,j) such that islands i and j are neighbours in the considered graph (we shall write $i\sim j$), a proportion $m_{i,j}$ of the individuals in island i migrate to island j at each generation (it is assumed that $\sum_{j:i\sim j} m_{i,j} = \sum_{j:i\sim j} m_{j,i}$). Letting the size of the subpopulation in each island go to infinity, the proportion of type 1 individuals in any given island converges to a diffusion similar to the Wright-Fisher diffusion but which presents a driving term (or a drift) taking into account the effect of migration ; this process is known as the Kimura stepping stone model and was first described in [Kim53]. For more details on this model, see [Eth09].

The pain in the torus For some populations, island models seem only a natural way to describe the spatial structure, but not always. In some cases, one would prefer to handle a continuous geographical space. It is at first tempting to try and see if one can obtain such a continuous space as a scaling limit of a stepping stone model, by letting the mesh of the graph go to zero. But then it is not clear what graph one should choose, and whether this choice affects the limiting model or not. It is at least possible to obtain a scaling limit of the stepping stone model on $\mathbb{Z}$ as a solution to a stochastic partial differential equation and the dual process is then a system of coalescing Brownian motions. Unfortunately, this equation has no solution in higher dimension, and the limit that one obtains when using the same scaling for the stepping stone model on $\mathbb{Z}^d$ is deterministic. In addition, note that two Brownian motions in higher dimension simply never meet, unlike one dimensional Brownian motion.

Thus modeling evolution in a spatial continuum is not as simple as it would appear, but it goes even deeper than that. In the 1940s, Wright and Malécot studied a model for populations evolving in a spatial continuum. In this model, at each generation, every individual produces a unit mean Poisson number of offspring whose locations are sampled according to a Gaussian distribution centered at the parent's location. Assuming that the population was distributed according to its stationary distribution, they were able to derive an approximate formula for the generating function of the time to the most recent common ancestor of two individuals, given their current location. However, as Felsenstein noted in [Fel75], this process has no stationary distribution because it evolves following a spatial branching process. Instead, the population will present clumps of high density. To overcome this, the population size has to be checked. Felsenstein tried to consider populations evolving in a torus instead of $\mathbb{R}^2$ and conditionned on the population size being constant, but this did not avoid clumping as he intended ; he called this problem *the pain in the torus.* The solution to avoid clumping is to regulate the population size locally. This can be done in a number of ways, but one of the most promising makes use of Poisson point processes.

Poisson point processes Point processes are a very general framework to study random countable sets. During WWII and the flying-bomb attack on London, the question was asked whether the German bombs were aimed at precise targets or if they were simply falling uniformly on the city. To answer this, one needs a model for this uniform configuration of impacts to compare it to the data. Poisson point processes provide just this model : random configurations of points. A point process on some measurable space E is a random measure Π on E such that, almost surely, there exists a countable set of points of E, $(x_i)_{i\in I}$ such that $\Pi = \sum_{i\in I} \delta_{x_i}$. Poisson point processes are a specific class of point processes whose definition is the following.

Definition 3.1 (Poisson point processes)**.** *Let $(E, \mathcal{A}, \mu)$ be a σ-finite measured space and Π be a point process on E. We say that Π is a Poisson point process based on μ if the following hold :*

i) $\forall A \in \mathcal{A}$, $\Pi(A)$ is a Poisson random variable with parameter $\mu(A)$,

ii) for any disjoint measurable sets $A_1, \dots, A_n$, the random variables $\Pi(A_1), \dots, \Pi(A_n)$ are independent.

Equivalently, for all disjoint measurable sets $A_1, \dots, A_n$,

$$\mathbb{P}\left(\Pi(A_1) = k_1, \dots, \Pi(A_n) = k_n\right) = \prod_{i=1}^{n} e^{-\mu(A_i)} \frac{\mu(A_i)^{k_i}}{k_i!}.$$

The next property gives a better idea of the shape of the set of points of Π.

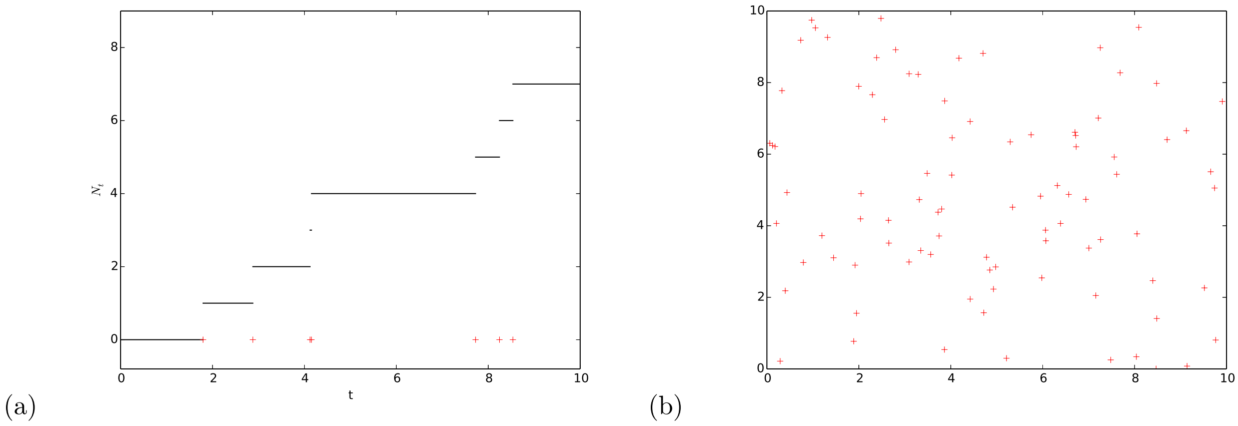


Figure 4: Poisson point processes in one and two dimensions.
In (a), each cross is a point of a Poisson point process on $\mathbb{R}_+$ based on Lebesgue measure and the counting process N_t is shown on vertical coordinates. Next (b) is a Poisson point process on $\mathbb{R}^2$ based on Lebesgue measure.

Proposition 3.2. *Assume that $\mu(E) < \infty$. Conditionally on the fact that $\Pi(E) = N$, Π has the same law as $\sum_{i=1}^{N} \delta_{x_i}$, where the x_i are independent random variables distributed according to $\frac{\mu(\cdot)}{\mu(E)}$.*

To see why this is true, suppose that $A_1, \dots, A_n$ is a partition of E and write, for $k_1 + \dots + k_n = N$:

$$\begin{aligned}\mathbb{P}\left(\Pi(A_1) = k_1, \dots, \Pi(A_n) = k_n \mid \Pi(E) = N\right) &= \prod_{i=1}^{n} e^{-\mu(A_i)} \frac{\mu(A_i)^{k_i}}{k_i!} \times \left(e^{-\mu(E)} \frac{\mu(E)^N}{N!}\right)^{-1} \\ &= \frac{N!}{k_1! \dots k_n!} \left(\frac{\mu(A_1)}{\mu(E)}\right)^{k_1} \dots \left(\frac{\mu(A_n)}{\mu(E)}\right)^{k_n}.\end{aligned}$$

This is particularly useful as it provides a way of constructing Poisson processes very easily, and thus to prove that they exist. Indeed, if E is of finite μ measure, then Proposition 3.2 gives an explicit construction of Π : take N to be a Poisson random variable with parameter $\mu(E)$ and sample $x_1, \dots, x_N$ independently according to $\frac{\mu(\cdot)}{\mu(E)}$,

$\Pi = \sum_{i=1}^{N} \delta_{x_i}$ is then a Poisson point process on E. If E is σ-finite, there exist disjoint subsets $(E_n)_{n\in\mathbb{N}}$ of finite μ measure whose union is E. On each of these sets, define a Poisson process Π_n based on the restriction of μ to E_n. Then $\Pi = \sum_n \Pi_n$ is a Poisson point process on E (we have used the Superposition Theorem, which says that a countable sum of Poisson processes is a Poisson process based on the sum of the measures on which the former are based, see [Kin92]).

Take for example $E = \mathbb{R}_+$ and suppose that Π is a Poisson point process on $\mathbb{R}_+$ based on Lebesgue measure. Set $N_t = \Pi([0,t))$. Then $(N_t)_{t\geq 0}$ is a Markov process with independent and identically distributed increments (*i.e.* it is a Lévy process). What's more N_t is a Poisson random variable with parameter t, thus $(N_t)_{t\geq 0}$ is in fact a Poisson process with intensity 1. One can also consider Poisson point processes based on Lebesgue measure on more general spaces. For instance, coming back to the London bombings, the spatial distribution of impacts was compared to a Poisson point process on $\mathbb{R}^2$ based on Lebesgue measure, yielding a very close fit and thus showing that the Germans were not aiming at any specific targets.

We have barely scratched the surface of the theory of Poisson point processes, [Kin92] is a much thorough though still accessible introduction to this topic.

The spatial Λ-Fleming-Viot process We are now able to present a model introduced in [BEV10] and [BEV13] which overcomes the pain in the torus by regulating the population size locally through the use of Poisson point processes. We shall start by defining a prelimiting model keeping track of the evolution of a countable but locally finite set of individuals and then construct the spatial Λ-Fleming-Viot process as the corresponding limiting process.

Let $\lambda > 0$. Suppose that we start with a set of individuals distributed on $\mathbb{R}^d$ according to a Poisson point process based on the measure $\lambda \mathrm{d}x$, each carrying one of two possible alleles of a given gene, 0 or 1. Let $\mu(\mathrm{d}r)$ and $\nu_r(\mathrm{d}u)$ be non trivial measures on $[0,\infty)$ and $(0,1]$, respectively. We assume

$$\int_0^\infty \int_0^1 u r^d \nu_r(\mathrm{d}u)\mu(\mathrm{d}r) < \infty. \tag{3.1}$$

Let Π be a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}^d \times [0,\infty) \times (0,1]$ based on $\mathrm{d}t \otimes \mathrm{d}x \otimes \mu(\mathrm{d}r)\nu_r(\mathrm{d}u)$. At each time t such that (t,x,r,u) is a point of Π, a reproduction event will take place within the ball of center x and radius r, u being called the impact parameter. The crucial point here is that the rate of these reproduction events does not depend on the state of the population, and (3.1) ensures that events affecting any given location x (*i.e.* x lies in the ball where the event takes place) will be locally finite. Let us precise what happens when a reproduction event of impact parameter u takes place in $B(x,r)$:

- If at time t the ball $B(x,r)$ is empty, nothing happens.
- If not, each individual in $B(x,r)$ dies with probability u, independently of each other.
- New individuals appear, distributed according to a Poisson point process based on the measure $u\lambda \mathbb{1}_{z\in B(x,r)}\mathrm{d}z$. Their parent is chosen uniformly at random among all individuals present in the ball just before time t, and they all inherit its type.

In [BEH09], it is proved that if λ is small, the population becomes extinct as $t \to \infty$ while for sufficiently large λ, it survives with high probability. We are interested in the limiting behaviour of this population when $\lambda \to \infty$, the size of the population being in units of size λ (*i.e.* each individual's contribution to the mass of the population is $\frac{1}{\lambda}$). To this end, let us define a measure m_t^λ on $\mathbb{R}^d \times \{0,1\}$ for all $t \geq 0$ by

$$m_t^\lambda(\mathrm{d}x\mathrm{d}k) = \frac{1}{\lambda}\sum \delta_{(x_i,k_i)}(\mathrm{d}x\mathrm{d}k)$$

where the sum is over all the individuals in the population at time t and x_i and k_i are respectively the location and the type of each individual. Note that $\left(m_t^\lambda\right)_{t\geq 0}$ is a (measure valued) process containing all the information about the successive states of the population. Specifically, $\int_{A\times\{1\}} m_t^\lambda(\mathrm{d}x\mathrm{d}k)$ is the number of type 1 individuals in the region A. The spatial Λ-Fleming-Viot process is defined as the limit of this process as $\lambda \to \infty$, we will denote it $(m_t)_{t\geq 0}$. We shall not precise the type of convergence that we consider, and we will be content with hinting at the arguments for the proof.

The process $(m_t)_{t\geq 0}$ is characterised as follows. For all $t \geq 0$, m_t is absolutely continuous with respect to Lebesgue measure on the spatial coordinate. Furthermore, since the number of individuals in any given domain A

is Poisson with parameter $\lambda |A|$ - where $|A|$ denotes Lebesgue measure of A - we will have that $m_t(A \times \{0,1\}) = |A|$. As a consequence, there exists a map $p_t : \mathbb{R}^d \to [0,1]$ (which is defined up to a negligible set) such that

$$m_t(\mathrm{d}x\mathrm{d}k) = \mathrm{d}x \otimes \mathcal{B}_{p_t(x)}(\mathrm{d}k)$$

where $\mathcal{B}_p$ is a probability measure on $\{0,1\}$ giving mass p to $\{1\}$ (*i.e.* it is the law of a Bernoulli random variable with parameter p). Note that

$$\int_{A\times\{1\}} m_t(\mathrm{d}x\mathrm{d}k) = \int_A p_t(x)\mathrm{d}x,$$

so that $p_t(x)$ can be seen as the proportion of type 1 individuals at location $x \in \mathbb{R}^d$ at time t. The process $(p_t)_{t\geq 0}$ evolves according to reproduction events driven by the same Poisson point process Π as the prelimiting model (the convergence can in fact be stated conditionally on Π, see [VW12]). At each reproducton event $(t,x,r,u) \in \Pi$, p_t is updated within the ball $B(x,r)$ in such a way that it reflects the average behaviour of m^λ over an infinite number of individuals :

- Choose a location y within $B(x,r)$ uniformly at random, and choose a type k such that $k = 1$ with probability $p_{t^-}(y)$.
- For all $z \in B(x,r)$, set
$$p_t(z) = (1-u)p_{t^-}(z) + u\mathbb{1}_{k=1},$$
and leave $p_t(z)$ unchanged for z outside of $B(x,r)$.

For the same reason as before, the rate at which reproduction events affect any given location $x \in \mathbb{R}^d$ is uniformly bounded by (3.1). In [BEV10], it is proved that this condition (3.1) is enough to ensure that a process evolving according to the above described dynamic exists and is unique (in law). The proof makes use of a powerful fact about duality, namely that existence of a dual process for a sufficiently wide class of functions implies uniqueness of the forwards in time process. Obviously, the duality relations for the spatial Λ-Fleming-Viot process are not as simple as for the Kingman coalescent and we shall simply describe the dual process and justify informally the link with the forwards in time process.

Spatial Λ-coalescent We call $\mathcal{A} = \{(A_1,x_1),\ldots,(A_k,x_k)\}$ a labelled partition of $\{1,\ldots,n\}$ if $\{A_1,\ldots,A_k\}$ is a partition of $\{1,\ldots,n\}$ and $x_1,\ldots,x_k$ are points in $\mathbb{R}^d$. We denote by $\mathcal{P}_n^l$ the set of labelled partitions of $\{1,\ldots,n\}$. The genealogy of n individuals sampled at locations $(x_1,\ldots,x_n)$ is described by a $\mathcal{P}_n^l$ valued process $(\mathcal{A}_t)_{t\geq 0}$ starting at the trivial partition $\{(\{1\},x_1),\ldots,(\{n\},x_n)\}$. For all $t \geq 0$, the blocks of $\mathcal{A}_t$ are thought of as containing the labels of individuals in the sample sharing a common ancestor at time t in the past, exactly as before, while the label of each block is the location of this ancestor.

As for the forwards in time process, let Π be a Poisson point process on $\mathbb{R}_+ \times \mathbb{R}^d \times [0,\infty) \times (0,1]$ based on the measure $\mathrm{d}t \otimes \mathrm{d}x \otimes \mu(\mathrm{d}r)\nu_r(\mathrm{d}u)$. At each point (t,x,r,u) of Π, each lineage (A_i,x_i) in $\mathcal{A}_{t^-}$ such that $x_i \in B(x,r)$ is affected independently with probability u. A location y is chosen uniformly at random in $B(x,r)$ and all the affected lineages merge into a single block with label y.

Condition (3.1) ensures that $(\mathcal{A}_t)_{t\geq 0}$ is a well defined Markov process with finite coalescence rates. It seems only reasonable that the coalescing events of $(\mathcal{A}_t)_{t\geq 0}$ should be driven by the same Poisson point process as the reproduction events of the spatial Λ-Fleming-Viot process. The proportion u of coalescing lineages in a given event corresponds to the offspring produced by the parent in the forwards in time process. Note that, contrary to the Kingman coalescent, more that two lineages can coalesce at a single time (in fact, $(\mathcal{A}_t)_{t\geq 0}$ can be seen as a spatial version of a Λ-coalescent, a class of coalescents introduced by Donnelly & Kurtz, Pitman and Sagitov). For a precise statement of the duality relations between $(m_t)_{t\geq 0}$ and $(\mathcal{A}_t)_{t\geq 0}$, see [BEV10] or [VW12] for a detailed construction of the dual process.

Rescaling limits of the spatial Λ-Fleming-Viot process Much research is currently being done within this framework for modeling evolution in a spatial continuum. Let us highlight one topic, namely finding rescaling limits of the spatial Λ-Fleming-Viot process. For instance, in [EVY14], the regime where the measure $\nu_r(\mathrm{d}u)$ concentrates around 0 is studied, more precisely, it is assumed that $\nu(\mathrm{d}u) = \delta_{u_n}(\mathrm{d}u)$ and that $u_n \to 0$. From a biological point of view, this corresponds to the behaviour of a population with infinite local density. Though it would seem from our construction of the spatial Λ-Fleming-Viot process that it is already dealing with a population with infinite size, it has nonetheless some feature of finite local population size since the offspring produced at each reproduction event forms a non-trivial proportion - namely u_n - of the population within the affected area. Thus letting u_n tend to zero amounts to letting the local population size go to infinity, and there is hope that it might yield a limit for $(p_t)_{t\geq 0}$ that is continuous in time, provided the right time scale is considered.

At the same time, two spatial scalings are studied. The first case is that of fixed radius reproduction events, where $\mu(\mathrm{d}r) = \delta_{r_n}(\mathrm{d}r)$, and we let $r_n \to 0$. This means that we assume that offspring remain close to their parent instead of spreading far away. This assumption of small offspring dispersal was used by R. A. Fisher in [Fis37] to motivate the use of a partial differential equation to model the evolution of a selectively advantageous mutation in a spatially distributed population. If $p(t, x)$ stands for the proportion of the mutant allele at $x \in \mathbb{R}^d$ and at time $t \geq 0$ and $s > 0$ is the intensity of the selection in favour of this mutation, Fisher asserts that p satisfies :

$$\frac{\partial p}{\partial t} = \Delta p + sp(1-p). \tag{3.2}$$

This equation was named the Fisher KPP equation, after Fisher, Kolmogorov, Petrovskii and Piscounov. We haven't specified how selection can be taken into account in the setting of the spatial Λ-Fleming-Viot process, but it is not difficult, and it is done in [EVY14]. The limit obtained for $(p_t)_{t\geq 0}$ for $d \geq 2$ when both r_n and u_n tend to zero, and with a suitable time scale, is precisely a solution to the Fisher KPP equation, thus providing a more mechanistic justification for the use of (3.2).

The second case involves reproduction events with arbitrarily large radius. Setting

$$\mu(\mathrm{d}r) = \frac{\mathbb{1}_{r\geq r_n}}{r^{d+\alpha+1}}\mathrm{d}r$$

for $\alpha \in (1, 2)$, we are interested in the limit when r_n (and u_n as well) tends to zero. It is shown that a deterministic limit still holds for $d \geq 2$, and that it is similar to the Fisher KPP equation, except that the Laplacian has to be replaced by a non local operator.

Further research (and that was the subject of my master's thesis) include obtaining stronger convergence results and studying the asymptotical fluctuations around the deterministic limit. It appears that the fluctuations are given by the solution to a stochastic partial differential equation driven by a Gaussian noise. We cannot introduce the subject of SPDE's here as it would require a substantial amount of technical results and preliminary knowledge about Itô's integral. Suffice it to say that satisfying results have been established in the neutral case, but that the proof so far fails to extend to the selective case, which could be the subject of future research.

References

[BEH09] N. Berestycki, A. Etheridge, and M. Hutzenthaler. Survival, extinction and ergodicity in a spatially continuous population model. In *Markov Proc. Rel. Fields*, 2009.

[BEV10] Nick H. Barton, Alison M. Etheridge, and Amandine Véber. A new model for evolution in a spatial continuum. *Electron. J. Probab*, 15(7), 2010.

[BEV13] N. H. Barton, A. M. Etheridge, and A. Véber. Modelling evolution in a spatial continuum. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):P01002, 2013.

[DK96] Peter Donnelly and Thomas G. Kurtz. A countable representation of the fleming-viot measure-valued diffusion. *The Annals of Probability*, 24(2):698–742, 1996.

[EK09] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

[Eth09] Alison Etheridge. Some mathematical models from population genetics. *Ecole d'Eté de Probabilités de Saint-Flour XXXIX—2009. To appear*, 2009.

[EVY14] Alison Etheridge, Amandine Veber, and Feng Yu. Rescaling limits of the spatial lambda-fleming-viot process with selection. *arXiv preprint arXiv:1406.5884*, 2014.

[Fel75] Joseph Felsenstein. A pain in the torus: some difficulties with models of isolation by distance. *American Naturalist*, pages 359–368, 1975.

[Fis37] Ronald Aylmer Fisher. The wave of advance of advantageous genes. *Annals of Eugenics*, 7(4):355–369, 1937.

[Kim53] M. Kimura. Stepping-stone model of population. *Annual Report of the National Institute of Genetics*, 3:62–63, 1953.

[Kin92] John Frank Charles Kingman. *Poisson processes*, volume 3. Oxford university press, 1992.

[Pit06] Jim Pitman. *Combinatorial stochastic processes*, volume 32. Springer, 2006.

[VW12] Amandine Veber and Anton Wakolbinger. The spatial lambda-fleming-viot process: an event-based construction and a lookdown representation. *arXiv preprint arXiv:1212.5909*, 2012.

[Wri43] Sewall Wright. Isolation by distance. *Genetics*, 28(2):114, 1943.