The spatial Lambda-Fleming-Viot process with fluctuating selection

Niloy Biswas
Department of Statistics
Harvard University
Science Center 400 Suite
One Oxford Street
Cambridge, MA 02138-2901
USA
niloy_biswas@g.harvard.edu

Alison M. Etheridge
Department of Statistics
Oxford University
24–29 St Giles
Oxford OX1 3LB
UK
etheridg@stats.ox.ac.uk

Aleksander Klimek *
Max Planck Institute
for Mathematics in Natural Sciences
Inselstrasse 22
04103 Leipzig
DE
klimek@mis.mpg.de

May 30, 2020

Abstract

We are interested in populations in which the fitness of different genetic types fluctuates in time and space, driven by temporal and spatial fluctuations in the environment. For simplicity, our population is assumed to be composed of just two genetic types. Short bursts of selection acting in opposing directions drive to maintain both types at intermediate frequencies, while the fluctuations due to 'genetic drift' work to eliminate variation in the population.

We consider first a population with no spatial structure, modelled by an adaptation of the Lambda (or generalised) Fleming-Viot process, and derive a stochastic differential equation as a scaling limit. This amounts to a limit result for a Lambda-Fleming-Viot process in a rapidly fluctuating random environment. We then extend to a population that is distributed across a spatial continuum, which we model through a modification of the spatial Lambda-Fleming-Viot process with selection. In this setting we show that the scaling limit is a stochastic partial differential equation. As is usual with spatially distributed populations, in dimensions greater than one, the 'genetic drift' disappears in the scaling limit, but here we retain some stochasticity due to the fluctuations in the environment, resulting in a stochastic p.d.e. driven by a noise that is white in time but coloured in space.

We discuss the (rather limited) situations under which there is a duality with a system of branching and annihilating particles. We also write down a system of equations that captures the frequency of descendants of particular subsets of the population and use this same idea of 'tracers', which we learned from [23] and [13], in numerical experiments with a closely related model based on the classical Moran model.

Key words: Spatial Lambda Fleming-Viot model, Fluctuating selection, stochastic growth models, Tracer dynamics, scaling limits

MSC 2010 **Subject Classification:** Primary: 60G57, 60J25, 92D15
Secondary: 60J75, 60G55

Contents

Contents

1 Introduction

A fundamental challenge in population genetics is to understand the balance between adaptive processes (*selection*) and random neutral processes (*genetic drift*). The most studied example of adaptation is directional selection acting on a single genetic locus. In the simplest model, each individual is either of type a or type A at the locus under selection, and the relative fitnesses of individuals carrying the two types is $1 + s_0 : 1$, for some small parameter s_0. At least provided that random fluctuations don't eliminate the favoured type before it can become established, natural selection will act to remove variability from the population until, in the absence of mutation, everyone is of the favoured type. However, there are other forms of selection that act to maintain genetic variation. In this paper we are concerned with populations that are subject to changing environmental conditions, that cause relative fitnesses of different genotypes to fluctuate in time and space. To quote *GILLESPIE* (2004, [20]), *"If fitnesses do depend on the state of the environment, as they surely must, then they must just as assuredly change in both time and space, driven by temporal and spatial fluctuations in the environment."*

We shall suppose that our population occurs in just two types (*alleles*), $\{a, A\}$ and that the environment fluctuates between two states, in the first of which a, and in the second of which A, is favoured. We suppose that selection is sufficiently strong that if the environment did not fluctuate,

2

the favoured type would rapidly fix in the population, but that there is a 'balance' between the two environments so that both types can be maintained at non-trivial frequencies for long periods of time. If the population has no spatial structure, then over large timescales the frequency of a-alleles can be modelled by a stochastic differential equation:

$$\mathrm{d}p = \mathbf{s}^2 p(1-p)(1-2p)\mathrm{d}t + \sqrt{p(1-p)}\mathrm{d}B^1 + \sqrt{2}\mathbf{s}p(1-p)\mathrm{d}B^2, \qquad (1.1)$$

where B^1 and B^2 are independent Brownian motions, the first (as we shall explain in Section 3) capturing the randomness due to genetic drift (that is the randomness due to reproduction in a finite population), the second encoding the random fluctuations in the environment (which are assumed to happen quickly on evolutionary timescales). The constant $\mathbf{s}$ is a scaled selection coefficient (see Section 3). For a population distributed across a one-dimensional spatial continuum, one can write down an analogous stochastic partial differential equation:

$$\mathrm{d}w(t,x) = \frac{1}{2}\Delta w(t,x)\mathrm{d}t + \mathbf{s}^2 w(t,x)(1-w(t,x))(1-2w(t,x))\mathrm{d}t$$
$$+ \sqrt{w(t,x)(1-w(t,x))}\mathcal{W}(\mathrm{d}t,\mathrm{d}x) + \sqrt{2}\mathbf{s}w(t,x)(1-w(t,x))W(\mathrm{d}t,\mathrm{d}x), \quad (1.2)$$

where $\mathcal{W}$ is space-time white noise (capturing genetic drift) and the independent noise W is white in time, but may be coloured in space reflecting spatial correlations in the environmental fluctuations. In the biologically most relevant case of two dimensions, this equation has no solution, and we must find a different approach.

The difficulties with modelling genetic drift in populations evolving in higher dimensional spatial continua, often referred to as 'the pain in the torus', are well known; see *BARTON, ETHERIDGE and VÉBER* (2013, [4]) for a review. They can be overcome using the spatial Lambda-Fleming-Viot process, introduced in *ETHERIDGE* (2008, [14]) and rigorously constructed in *BARTON, ETHERIDGE and VÉBER* (2010, [3]), and here we adapt that model to incorporate fluctuating selection.

Our first result deals with the non-spatial case. We take a scaling limit of the Lambda-Fleming-Viot process and recover (1.1), which coincides with that obtained by *GILLESPIE* (2004, [20]) as a scaling limit of a Wright-Fisher type model. We then turn to the scaling limit of the spatial Lambda-Fleming-Viot process with fluctuating selection. In dimension one, the limiting process coincides (up to constants) with the stochastic p.d.e. (1.2). In higher dimensions, the term corresponding to genetic drift vanishes in the limit, but the effects of the fluctuations in the environment can still persist, resulting in a stochastic p.d.e. driven by (spatially) coloured noise.

Our ultimate aim is to find ways to distinguish the effects of spatial and temporal environmental fluctuations on genetic data. This would involve understanding the genealogical trees relating individuals in a sample from the population. Although for our prelimiting model we can write down an analogue of the ancestral selection graph of *KRONE and NEUHAUSER* (1997, [27]), *NEUHAUSER and KRONE* (1997, [37]), which tracks all 'potential ancestors' of individuals in a sample from the population, this process seems to be rather unwieldy. Moreover, when we apply our rescaling, the scaled ancestral selection graphs do not converge and we have not found a satisfactory way to extract genealogies for the limiting model. When selection does not fluctuate, the original ancestral selection graph can be thought of as a moment dual to the forwards in time diffusion describing allele frequencies in the population. It is natural to ask whether there are other dual processes that we could exploit when selection fluctuates. Our attempts to find a useful dual for the equation (1.2) have met with limited success, but in Section 5 we show that (after an affine transformation) there are circumstances in which a branching and annihilating dual exists.

In the absence of a useful dual process, instead we take a first step towards understanding ancestry in the population by following an interesting approach of *HALLATSCHEK and NELSON* (2008, [23]) and, more recently, *DURRETT and FAN* (2016, [13]) which uses the idea of 'tracers' to explore the way in which descendants of a subpopulation of the type a individuals evolve forwards in time. In Section 6 we write down the system of stochastic p.d.e.'s that will determine the tracer dynamics. This idea is exploited further in our numerical experiments of Section 7.

The rest of the article is laid out as follows. In Section 2 we very briefly outline some of the biological background. In Section 3 we consider the case in which the population has no spatial structure. To prepare the ground for the case of spatially structured populations, we work with the Lambda-Fleming-Viot process (also sometimes known as the generalised Fleming-Viot process) that was introduced in *DONNELLY and KURTZ* (1999, [12]), *BERTOIN and LE GALL* (2003, [6]). In particular, we investigate different scaling limits, reflecting longtime behaviour of the process for different balances between the rate of changes of environment and the strength of selection. In Section 4 we define the spatial Lambda-Fleming-Viot process with fluctuating selection and give a precise statement of our scaling result for this model. In Section 5, we discuss the situations in which we can investigate the limiting process through duality with a system of branching and annihilating particles. Tracers are introduced in Section 6 and then explored numerically (for a Moran model of a subdivided population) in Section 7. The proof of our main scaling limit is in Section 8. The appendices contain some (important) technical results that we require in the course of the proofs.

2　　Biological background

In this section we outline the biological context for this work. Although not a prerequisite for understanding the mathematics of subsequent sections, it explains our motivation for tackling this particular scaling limit.

Suppose that a gene occurs in just two forms that, because of environmental fluctuations, each finds itself subject to short alternating bursts of positive and negative selection. Even if these changes are happening on a much faster scale than neutral evolution, they may influence gene frequencies. For example, in diploid individuals (carrying two copies of the gene), a heterozygote (carrying one allele of each type) may have higher mean fitness, when we take account of different environments, than either homozygote, and so allelic variation can be maintained for long periods, even though at any given time the population is subject to directional selection. This *marginal overdominance* is an example of *balancing selection*. In equation (1.1) we see this in the deterministic term $(sp(1-p)(1-2p)\mathrm{d}t)$ on the right hand side. *WRIGHT* (1969, [49]) observed that spatial heterogeneity in the direction of selection combined with density dependent reproduction can also lead to *balanced polymorphism*, that is adaptive alleles are held at intermediate frequencies for long periods (see also *DELPH and KELLY* (2014, [10]) and references therein).

One of the recurring arguments in evolutionary biology is whether evolution occurs principally through natural selection or through neutral processes, in which no particular genetic type is favoured, such as genetic drift. A data set that has sat at the heart of this debate for the last 70 years is a time series of changes in the genotype frequency of a polymorphism of the Scarlet Tiger Moth, *Callimporpha (Panaxia) dominula*, in an isolated population at Cothill Fen near Oxford,

UK. *FISHER and FORD* (1947, [17]) found that the proportion of a certain *medionigra* allele in the population increased significantly between 1929 and 1941 from 1.2% to 11.1%, and decreased to 5.2% between 1941 and 1946. They concluded, "... the observed fluctuations in generations are much greater than could be ascribed to random survival only. Fluctuations in natural selection must therefore be responsible for them.". Fisher (a strong proponent of the importance of selection) was challenged by Wright (a champion of genetic drift) who argued that multiple factors could be at play and, moreover, Fisher may have underestimated the strength of genetic drift. *O'HARA 2005* (2005, [40]) analysed the, by then 60 year long, time series of data from Cothill and concluded that most of the pattern of variation in the population should be attributed to genetic drift. Moreover, although selection is acting, mean fitness barely increased.

It is unusual to have such a long time series of data, especially in conjunction with information about the environment. In general it will also be far from clear which genes are undergoing selection, rather one tries to infer the action of selection through studying neutral diversity. For most populations, it may be very difficult to distinguish fluctuating selection from genetic drift. To see why, we recall a model due to Gillespie that captures the effect of a series of 'selective sweeps' through a population. Suppose that a selectively favoured mutation arises at some point on the genome and rapidly increases in frequency (until the whole population carries it). Because genes are arranged on chromosomes, different genes do not evolve independently of one another. As a result of a process called recombination, correlations between genes decrease as a function of the distance between them on the chromosome. Nonetheless, a neutral allele fortunate enough to be on the same chromosome as the selectively favoured mutation will itself receive a boost in its frequency (even if as a result of recombination it doesn't exhaust the whole population). This boost to the type at the neutral locus is known as 'genetic hitchhiking', a term introduced by *MAYNARD SMITH and HAIGH* (1974 [34]). Of course, correspondingly, a neutral allele associated with an unfavoured type will decrease in frequency. *GILLESPIE* (2000, [18]), *GILLESPIE* (2001, [19]) investigated a model in which strongly selected mutations which give rise to hitchhiking events occur at the points of a Poisson process. He assumes that selection is strong enough that the duration of the sweeps causing the hitchhiking events that affect a given locus is small compared to the time between them so that we can ignore the possibility that a locus will be subject to two simultaneous hitchhiking events. He establishes that the first two moments of the change in allele frequency at the neutral locus over the course of a hitchhiking event take exactly the same form as if they had been produced by genetic drift over a single generation of reproduction. It is not hard to see that we will see the same hitchhiking effects under 'partial sweeps' driven by environmental fluctuations. This process of 'genetic draft' induced by selection, strongly resembles genetic drift and it may be hard to distinguish the two. *BARTON* (2000, [2]) also considers the 'genetic drift' induced by hitchhiking.

Not surprisingly, the impact of environmental fluctuations on genetic variation has been extensively studied. Nonetheless, even in the absence of spatial structure, it remains an open question to characterise situations under which fluctuating environmental conditions can maintain genetic variation; see e.g. *NOVAK and BARTON* (2017, [39]). Moreover, the effects of genetic drift have been largely ignored. This is perhaps because acting in isolation, genetic drift typically impacts gene frequencies over periods of (tens of) thousands of generations, much longer than the time scales of climatic fluctuation. However, once a particular genotype becomes rare, perhaps as a result of a run of unfavourable environments, stochastic fluctuations will be dominated by genetic drift, through which the genotype can be lost.

A further challenge in identifying genes that are subject to fluctuating selection is that, even if we can disentangle the effects of drift, numerous selection schemes lead to forms of balancing selection. For example, in the absence of spatial structure, allele frequency dynamics under fluctuating

selection are identical to those under within-generation fecundity variance polymorphism. In this setting, *TAYLOR* (2013, [46]) shows that the effects on the genealogy at a linked neutral locus will differ. *FIJARCZYK and BABIK* (2015, [16]) and the references therein provide an overview of theoretical and empirical evidence for various forms of balancing selection and methods for their detection.

Recently, *BERGLAND, BEHRMAN, O'BRIEN, SCHMIDT and PETROV* (2014, [5]) reported hundreds of polymorphisms in *Drosophila melanogaster* whose frequencies oscillate among seasons and they attribute this to strong, temporally variable selection. They also cite evidence that genetic (and phenotypic) variation is maintained by temporally fluctuating selection for a variety of other organisms. *GOMPERT* (2016, [21]) proposes an approach to quantifying variable selection in populations experiencing both spatial and temporal variations in selection pressure. In spite of this combination of theoretical and empirical evidence for the importance of fluctuating selection, we have only a limited understanding of some basic questions: how many loci are subject to temporally fluctuating selection? How strong is that selection? What is the relationship between temporally and spatially varying selection?

Since natural environments are never truly constant, it is clearly important to understand the implication of temporally and spatially varying selection pressures. *CVIJOVIC, GOOD, JERISON and DESAI* (2015, [9]) examines some of the implications of temporal fluctuations. Our work here is a step towards a tractable framework in which to consider the combined effects of spatial and temporal fluctuations.

3 The non-spatial case

3.1 The (non-spatial) model

We first consider a population without spatial structure. Although we would obtain exactly the same scaling limits if we were to use the classical Moran or Wright-Fisher models as the basis of our approach, c.f. *GILLESPIE* (2004, [20]), for consistency with what follows, we shall work with the (non-spatial) Lambda-Fleming-Viot process. The key ideas that will be required in the spatial setting already appear here, where they are not obscured by notational complexity. An analogous scaling limit is obtained for the Wright-Fisher model with fluctuating selection (using similar reasoning) in *HUTZENTHALER, PFAFFELHUBER and PRINTZ* (2018, [25]).

We shall restrict ourselves to the special case of the Lambda-Fleming-Viot process in which reproduction events fall at a finite rate, determined by a Poisson process. We shall also suppose that there are just two types of individual, $\{a, A\}$. In each event, a parent is chosen from the population immediately before the event, and a portion u of the population is replaced by offspring of the same type as the parent. In general the quantity u, which we shall call the *impact* of the event, may be random. Selection (on fecundity) can be incorporated by weighting the choice of parent, to favour one type or the other, and we shall extend previous versions of the model to allow the direction of selection to fluctuate. More precisely, we have the following definition.

Definition 3.1 (Lambda-Fleming-Viot process with fluctuating selection). *The Lambda-Fleming-Viot process with fluctuating selection, $\{p(t)\}_{t \geq 0}$ is a càdlàg process taking its values in $[0, 1]$, with $p(t)$ to be interpreted as the proportion of type a individuals in the population at time t.*

Let Π be a Poisson process defined on $\mathbb{R} \times (0, 1) \times (0, 1)$ with intensity measure $\mathrm{d}t \otimes \nu(\mathrm{d}u) \otimes \sigma(\mathrm{d}s)$, where ν and σ are some probability measures. Moreover, let Π^{env} be a rate τ^{env} Poisson process, independent of Π (where $\tau^{env} \in (0, \infty)$). The state of the environment is a random variable $\zeta(t) \in \{-1, 1\}$. At the times of the Poisson process Π^{env}, ζ is resampled uniformly from $\{-1, 1\}$.

The dynamics of $\{p(t)\}_{t \geq 0}$ can be described as follows. If $(t, u, \mathbf{s}) \in \Pi$, a reproduction event occurs. Then:

1. *select a parental type $\kappa \in \{a, A\}$ according to*

$$\mathbb{P}[\kappa = a] = \frac{(1 + \mathbf{s})p(t_-)}{1 + \mathbf{s}p(t_-)} \quad if \quad \zeta = -1,$$

$$\mathbb{P}[\kappa = a] = \frac{p(t_-)}{1 + \mathbf{s}(1 - p(t_-))} \quad if \quad \zeta = 1.$$

2. *A proportion u of the population immediately before the event dies and is replaced by offspring of the chosen type, that is*

$$p(t) = (1 - u)p(t_-) + \mathbf{1}_{\{\kappa = a\}}u.$$

Remark 3.2. *If, instead of resampling the environment according to an independent Poisson process, we resampled it at each reproduction event, by choosing ν to be distributed as the proportion of hitchhikers when a selective sweep occurs at a random distance from our chosen locus, we would recover Gillespie's model of genetic draft at a neutral locus linked to loci undergoing a sequence of selective sweeps.*

We shall see that the rate of resampling of the environment (relative to the strength of selection in each event) plays a key role in the long term behaviour of the population.

3.2 Scaling limits

In order to simplify the notation still further, we specialise to the case in which the Poisson point process Π of Definition 3.1 has intensity $dt \otimes \delta_{\bar{u}} \otimes \delta_{\mathbf{s}}$ for some fixed $\bar{u}$ and $\mathbf{s}$; in other words we fix the impact and the strength of selection in each event. A general result can be obtained from our calculations below by integration.

In order to obtain a diffusion approximation, we shall speed up the rate of reproduction events by a factor n, but scale down both the impact and the strength of the selection. We write $u_n, \mathbf{s}_n$ for the impact and strength of selection at the nth stage of our rescaling. We shall also scale Π^{env} to have rate n^γ, with $\gamma > 0$ to be chosen. We shall need the joint generator $\mathcal{L}^{(n)}$ of the pair (p, ζ) at the nth stage of this rescaling. We write π for the uniform measure on $\{-1, 1\}$ and $\mathbb{E}_\pi$ for the corresponding expectation. In an obvious notation, for suitable test functions f, we have

$$\mathcal{L}_n f(p, \zeta) = \mathbf{1}_{\{\zeta = -1\}} n \left(\left[\frac{(1 + \mathbf{s}_n)p}{1 + \mathbf{s}_n p} \right] f((1 - u_n)p + u_n, \zeta) \right.$$

$$\left. + \left[\frac{1 - p}{1 + \mathbf{s}_n p} \right] f((1 - u_n)p, \zeta) - f(p, \zeta) \right)$$

$$+ \mathbf{1}_{\{\zeta = 1\}} n \left(\left[\frac{p}{1 + \mathbf{s}_n(1 - p)} \right] f((1 - u_n)p + u_n, \zeta) \right.$$

$$\left. + \left[\frac{(1 + \mathbf{s}_n)(1 - p)}{1 + \mathbf{s}_n(1 - p)} \right] f((1 - u_n)p, \zeta) - f(p, \zeta) \right)$$

$$+ n^\gamma \left(\mathbb{E}_\pi[f(p, \cdot)] - f(p, \zeta) \right).$$

Expanding the ratios involving $\mathbf{s}_n$ as geometric series, and using Taylor's Theorem to expand f (as a function of p), we obtain

$$
\begin{aligned}
\mathcal{L}_n f(p, \zeta) = {} & n\{p - \mathbf{s}_n \zeta p(1-p) + \mathcal{O}(\mathbf{s}_n^2)\} \\
& \times \{u_n(1-p)f'(p,\zeta) + \frac{1}{2}u_n^2(1-p)^2 f''(p,\zeta) + \mathcal{O}(u_n^3)\} \\
& + n\{(1-p) + \mathbf{s}_n \zeta p(1-p) + \mathcal{O}(\mathbf{s}_n^2)\} \\
& \times \{-u_n p f'(p,\zeta) + \frac{1}{2}u_n^2 p^2 f''(p,\zeta) + \mathcal{O}(u_n^3)\} \\
& + n^\gamma\left(\mathbb{E}_\pi[f(p,\cdot)] - f(p,\zeta)\right) \\
= {} & \frac{1}{2}n u_n^2 p(1-p)f''(p,\zeta) - n\mathbf{s}_n u_n \zeta p(1-p)f'(p,\zeta) \\
& + n^\gamma\left(\mathbb{E}_\pi[f(p,\zeta)] - f(p,\zeta)\right) + \mathcal{O}\left(n(\mathbf{s}_n^2 u_n + u_n^2 \mathbf{s}_n + u_n^3)\right).
\end{aligned}
\tag{3.1}
$$

In order to obtain a diffusion limit, we see that we should take nu_n^2 to be $\mathcal{O}(1)$. If the environment didn't change, then we would require $nu_n\mathbf{s}_n$ to be $\mathcal{O}(1)$ and on passage to the limit recover the classical Wright-Fisher diffusion with selection, whose generator, if $\zeta = -1$ say, takes the form

$$
\mathcal{L}^{WFS}f(p) = \frac{1}{2}p(1-p)f''(p) + \mathbf{s}p(1-p)f'(p).
$$

Since we are modelling short bursts of strong selection, we set

$$
u_n = n^{-1/2}\bar{u}, \quad \mathbf{s}_n = n^{-1/2+\alpha}\mathbf{s},
\tag{3.2}
$$

for some $\alpha \in (0, 1/4)$. The restriction $\alpha < 1/4$ ensures that the error term $n\mathbf{s}_n^2 u_n$ in the expression (3.1) is negligible as $n \to \infty$.

We can then write the rescaled generator in the form

$$
\mathcal{L}_n f(p,\zeta) = \mathcal{L}^{neu}f(p,\zeta) + n^\alpha \mathcal{L}^{fsel}f(p,\zeta) + n^\gamma \mathcal{L}^{env}f(p,\zeta) + \mathcal{O}(n^{-1/2+2\alpha}),
\tag{3.3}
$$

where

$$
\begin{aligned}
\mathcal{L}^{neu}f(p,\zeta) &= \frac{1}{2}\bar{u}^2 p(1-p)f''(p,\zeta) \\
\mathcal{L}^{fsel}f(p,\zeta) &= -\zeta\bar{u}\mathbf{s}p(1-p)f'(p,\zeta) \\
\mathcal{L}^{env}f(p,\zeta) &= \mathbb{E}_\pi[f(p,\zeta)] - f(p,\zeta).
\end{aligned}
$$

To see how we should choose γ, we employ a 'separation of timescales' trick due to KURTZ (1973, [28]). We apply the generator (3.3) to test functions of the form

$$
g(p,\zeta) = f(p) + n^{-\delta}\left(\mathcal{L}^{fsel}f\right)(p,\zeta).
\tag{3.4}
$$

For this choice, we obtain

$$
\begin{aligned}
\mathcal{L}_n g(p,\zeta) = {} & \mathcal{L}^{neu}f(p) + n^\alpha\left(\mathcal{L}^{fsel}f\right)(p,\zeta) + n^{-\delta}\mathcal{L}^{neu}\left(\mathcal{L}^{fsel}f\right)(p,\zeta) \\
& + n^{\alpha-\delta}\mathcal{L}^{fsel}\left(\mathcal{L}^{fsel}f\right)(p,\zeta) - n^{\gamma-\delta}\left(\mathcal{L}^{fsel}f\right)(p,\zeta) + \mathcal{O}(n^{-1/2+2\alpha}), \quad (3.5)
\end{aligned}
$$

where we have used the fact that $\mathcal{L}^{env}f(p) = 0$ (since f does not depend on ζ) and $\mathbb{E}_\pi[n^{-\delta}\left(\mathcal{L}^{fsel}f\right)(p,\zeta)] = 0$, since $\mathbb{E}_\pi[\zeta] = 0$.

8

Evidently, to obtain a non-trivial limit we should take $\gamma = \delta + \alpha$. The most interesting case is when $\delta = \alpha$ and so $\gamma = 2\alpha$. In that case, letting $n \to \infty$, in the limit the equation (3.5) becomes

$$\mathcal{L}f(p, \zeta) = \mathcal{L}^{neu}f(p) + \mathcal{L}^{fsel}\left(\mathcal{L}^{fsel}f\right)(p, \zeta). \tag{3.6}$$

To evaluate the right hand side,

$$\mathcal{L}^{fsel}\left(\mathcal{L}^{fsel}f\right)(p, \zeta) = \mathcal{L}^{fsel}\left(-\zeta\bar{u}\mathbf{s}p(1-p)f'(p)\right) = \zeta^2\bar{u}^2\mathbf{s}^2 p(1-p)\frac{d}{dp}\left(p(1-p)\frac{d}{dp}f(p)\right)$$

$$= \zeta^2\bar{u}^2\mathbf{s}^2 p(1-p)(1-2p)f'(p) + \zeta^2\bar{u}^2\mathbf{s}^2 p^2(1-p)^2 f''(p).$$

Noting that $\zeta^2 \equiv 1$, equation (3.6) then reads

$$\mathcal{L}f(p) = \bar{u}^2\mathbf{s}^2 p(1-p)(1-2p)f'(p) + \left(\frac{1}{2}\bar{u}^2 p(1-p) + \bar{u}^2\mathbf{s}^2 p^2(1-p)^2\right)f''(p).$$

Remark 3.3. *There are other limits that can be obtained when $\gamma > 2\alpha$. For example if $\alpha = 1/4$ and we resample the environment at every reproduction event, corresponding to $\gamma > 1$, then [35] shows that, under the same scaling of u_n, the frequency of type a alleles in the population converges weakly to the solution of*

$$\mathrm{d}p = \frac{1}{2}\mathbf{s}^2\bar{u}p(1-p)(1-2p)\mathrm{d}t + \bar{u}\sqrt{p(1-p)}\mathrm{d}B_t$$

for a standard Brownian motion $\{B_t\}_{t\geq 0}$. The deterministic drift here arises from the term of order $nu_n\mathbf{s}_n^2$ that under our previous scaling we were able to neglect in (3.1).

Based on these calculations, the following proposition follows easily from Theorem 2.1 of [29], which we recall later as Theorem 8.1. In the interests of space, we omit the details of the proof, which follows from exactly the same arguments as those that we employ in the spatial setting.

Proposition 3.4. *Let $\{p^{(n)}(t)\}_{t\geq 0}$ denote the (non-spatial) Lambda-Fleming-Viot process of Definition 3.1 in which Π has intensity $n\mathrm{d}t \otimes \delta_{u_n} \otimes \delta_{\mathbf{s}_n}$, where*

$$u_n = n^{-1/2}\bar{u}, \quad \mathbf{s}_n = n^{-1/2+\alpha}\mathbf{s}, \text{ and } \tau^{env} = n^{2\alpha},$$

for some $\alpha \in (0, 1/4)$. Suppose further that the sequence of initial conditions $\{p^{(n)}(0)\}_{n\geq 1}$ converges to p_0 as $n \to \infty$. Then as n tends to infinity, $\{p^{(n)}(t)\}_{t\geq 0}$ converges weakly in $\mathcal{D}([0,\infty), [0,1])$ (the space of càdlàg functions taking values in $[0,1]$) to the one-dimensional diffusion with drift

$$\bar{u}^2\mathbf{s}^2 p(1-p)(1-2p)$$

and quadratic variation

$$\bar{u}^2 p(1-p) + 2\bar{u}^2\mathbf{s}^2 p^2(1-p)^2,$$

started from p_0. In other words, the limiting process is the unique weak solution to the equation

$$\mathrm{d}p(t) = \bar{u}^2\mathbf{s}^2 p(t)(1-p(t))(1-2p(t))\mathrm{d}t$$
$$+ \bar{u}\sqrt{p(t)(1-p(t))}\mathrm{d}B_t^{(1)} + \sqrt{2}\bar{u}\mathbf{s}p(t)(1-p(t))\mathrm{d}B_t^{(2)}, \tag{3.7}$$

with $p(0) = p_0$, and where $\{B_t^{(1)}\}_{t\geq 0}$ and $\{B_t^{(2)}\}_{t\geq 0}$ are two independent standard Brownian motions.

4   Definition and scaling of the SLFVFS

In this section we first extend the Lambda-Fleming-Viot model with fluctuating selection of Section 3 to the spatial setting. The idea is simple: reproduction events are still driven by a Poisson point process, but now, in addition to specifying the strength of selection and the impact associated with each event, we must also specify the spatial region in which it takes place. As has become usual in this framework, we shall take those regions to be closed balls (indeed for simplicity we shall take our events to be of a fixed radius), but the same results will hold under much more general conditions, subject to some symmetry and boundedness assumptions. Having defined the model, we state our main scaling result for the spatial model.

4.1   Spatial Lambda-Fleming-Viot process with fluctuating selection

We suppose that the population, which is distributed across $\mathbb{R}^d$, is subdivided into two genetic types $\{a, A\}$. As explained in detail in *ETHERIDGE, VÉBER and YU* (2019, [15]), which in turn borrows results from *VÉBER and WAKOLBINGER* (2015, [48]), formally, at each time the state of the population is described by a measure M_t on $\mathbb{R}^d \times K$, where $K = \{a, A\}$, whose first marginal is Lebesgue measure on $\mathbb{R}^d$. At any fixed time there is a density $w(t, \cdot) : \mathbb{R}^d \to [0, 1]$ such that

$$M_t(dx, d\kappa) = \big(w(t, x)\delta_a(d\kappa) + (1 - w(t, x))\delta_A(d\kappa)\big)\mathrm{d}x.$$

Of course $w(t, x)$, which one should interpret as the proportion of the population at the location x at time t that is of type a, is only defined up to a Lebesgue null set. We shall topologize $\mathcal{M}_\lambda$ by declaring that w_n converges to w if for any bounded and compactly supported function f

$$\left(\int_{\mathbb{R}^d} f(x)w_n(x)dx, \int_{\mathbb{R}^d} f(x)w_n^2(x)dx, \int_{\mathbb{R}^d} f(x)w_n^3(x)dx\right)$$
$$\to \left(\int_{\mathbb{R}^d} f(x)w(x)dx, \int_{\mathbb{R}^d} f(x)w^2(x)dx, \int_{\mathbb{R}^d} f(x)w^3(x)dx\right).$$

The space $\mathcal{M}_\lambda$ with this topology is metrizable and compact, c.f. Section 2 of *VÉBER and WAKOLBINGER* (2015, [48]). A brief discussion of the topology can be found in Appendix C.

In what follows, we shall consider a representative of the density of M_t. It will be convenient to fix a representative $w(0, \cdot)$ of M_0 and then update it using the procedure described in the definition below, but the reader should bear in mind that the fundamental object is the measure-valued evolution. This becomes important when we talk about convergence of our rescaled processes; tightness will be immediate in the space of measures, but we will need to work harder to identify the dynamics of the density of the limit.

In what follows, for every $f \in C_c$ (continuous functions of compact support on $\mathbb{R}^d$) we shall use the notation

$$\langle w, f \rangle = \int_{\mathbb{R}^d} w(x)f(x)\mathrm{d}x.$$

Recall that the limit that we obtained in Section 3 corresponded to our throwing away the terms of order $n\mathbf{s}_n^2 u_n$ in (3.1). In other words we approximated $(1 + \mathbf{s})p/(1 + \mathbf{s}p)$ by $p + \mathbf{s}p(1 - p) = p(1 - \mathbf{s}) + \mathbf{s}\left(1 - (1 - p)^2\right)$ and similarly $p/(1 + \mathbf{s}(1 - p))$ was approximated by $p - \mathbf{s}p(1 - p) = p(1 - \mathbf{s}) + \mathbf{s}p^2$. Under this approximation, since reproduction events are based on a Poisson process of events, we

can think of splitting those events into two types: neutral events and selective events. In the non-spatial setting, neutral events occur at rate $(1 - \mathbf{s})$ and, for such an event, the chance that the parent is type a is p. Selective events fall at rate $\mathbf{s}$. One then selects two 'potential' parents. If $\zeta = -1$, then the offspring are type a provided not both potential parents are type A, which has probability $1 - (1 - p)^2$, whereas if $\zeta = 1$, the offspring are type a only if both potential parents are type a (probability p^2).

To avoid additional algebra, we shall define the spatial version of our model using this approximation. In our main scaling result, we shall indeed choose our scaling in such a way that $n\mathbf{s}_n^2 u_n \to 0$ as $n \to \infty$.

We shall assume that the random field which specifies the environment is of the form $2X\mathbf{1}_B - 1$, for a random closed subset $B \subseteq \mathbb{R}^d$ and X a Bernoulli random variable with parameter $1/2$. Our main result will require some additional regularity which guarantees that changes in the environment are typically seen over larger spatial scales than reproduction events. This is made precise in (4.3) below.

Definition 4.1 (Spatial Lambda-Fleming-Viot process with fluctuating selection (SLFVFS))**.** *Let μ be a measure on $(0, \infty)$ and for each $r \in (0, \infty)$, let ν_r be a probability measure on $[0, 1]$, such that the mapping $r \to \nu_r$ is measurable and*

$$\int_{(0,\infty)} r^d \int_{[0,1]} u \, \nu_r(\mathrm{d}u)\mu(\mathrm{d}r) < \infty. \tag{4.1}$$

Further, fix $\mathbf{s} \in [0, 1]$ and let Π^{neu}, Π^{fsel}, be independent Poisson point processes on $\mathbb{R}_+ \times \mathbb{R}^d \times (0, \infty) \times [0, 1]$ with intensity measures $(1 - \mathbf{s})\mathrm{d}t \otimes \mathrm{d}x \otimes \mu(\mathrm{d}r)\nu_r(\mathrm{d}u)$ and $\mathbf{s}\mathrm{d}t \otimes \mathrm{d}x \otimes \mu(\mathrm{d}r)\nu_r(\mathrm{d}u)$ respectively.

Let Π^{env} be a Poisson process, independent of Π^{neu}, Π^{fsel}, with intensity τ^{env}, dictating the times of the changes in the environment. Let $\{\xi^{(m)}(\cdot)\}_{m \geq 0}$ be a family of independent identically distributed random fields of the form $\xi^{(m)} = 2X^{(m)}\mathbf{1}_{B^{(m)}} - 1$, where $\{B^{(m)}\}_{m \geq 0}$ are random closed subsets of $\mathbb{R}^d$ and $\{X^{(m)}\}_{m \geq 0}$ are Bernoulli random variables with parameter $1/2$. We write

$$\mathbb{P}\left[\xi^{(m)}(x) = -1\right] = \frac{1}{2} = \mathbb{P}\left[\xi^{(m)}(x) = +1\right],$$
$$\mathbb{E}\left[\xi^{(m)}(x)\xi^{(m)}(y)\right] = g(x, y),$$

where the covariance function $g(x, y)$ will be assumed to be an element of $C_b\left(\mathbb{R}^d \times \mathbb{R}^d\right)$. Set $\tau_0 = 0$ and write $\{\tau_m\}_{m \geq 1}$ for the points in Π^{env} and define

$$\zeta(t, \cdot) := \sum_{m=0}^{\infty} \xi^{(m)}(\cdot)\mathbf{1}_{[\tau_m, \tau_{m+1})}(t).$$

In other words, the environment $\zeta(t, \cdot)$ is resampled, independently, at the times of the Poisson process Π^{env}.

The spatial Lambda-Fleming-Viot process with fluctuating selection (SLFVFS) *with driving noises Π^{neu}, Π^{fsel}, Π^{env}, is the $\mathcal{M}_\lambda$-valued process $\left(M_t\right)_{t \geq 0}$ with dynamics described as follows. Let $w(t_-, \cdot)$ be a representative of the density of M_{t_-} immediately before an event (t, x, r, u) from Π^{neu} or Π^{fsel}. Then the measure M_t immediately after the event has density $w(t, \cdot)$ determined by:*

1. *If $(t, x, r, u) \in \Pi^{neu}$, a neutral event occurs at time t within the closed ball $B(x, r)$. Then*

   (a) *Choose a parental location l according to the uniform distribution on $B(x, r)$.*

(b) *Choose the parental type $\kappa \in \{a, A\}$ according to the distribution*

$$\mathbb{P}\left[\kappa = a\right] = w(t_-, l), \quad \mathbb{P}\left[\kappa = A\right] = 1 - w(t_-, l).$$

(c) *A proportion u of the population within $B(x, r)$ dies and is replaced by offspring with type κ. Therefore, for each point $y \in B(x, r)$,*

$$w(t, y) = w(t_-, y)(1 - u) + u\mathbf{1}_{\{\kappa = a\}}.$$

2. *If $(t, x, r, u) \in \Pi^{fsel}$, a selective event occurs at time t within the closed ball $B(x, r)$. Then*

    (a) *Choose two parental locations l_0, l_1 independently, according to the uniform distribution on $B(x, r)$.*

    (b) *Choose the two parental types, κ_0, κ_1, independently, according to*

$$\mathbb{P}\left[\kappa_i = a\right] = w(t_-, l_i), \quad \mathbb{P}\left[\kappa_i = A\right] = 1 - w(t_-, l_i).$$

    (c) *A proportion u of the population within $B(x, r)$ dies and is replaced by offspring with type chosen as follows:*

        i. *If $\zeta(t, x) = 1$, their type is set to be a if $\kappa_0 = \kappa_1 = a$, and A otherwise. Thus, for each $y \in B(x, r)$,*

$$w(t, y) = (1 - u)w(t_-, y) + u\mathbf{1}_{\{\kappa_0 = \kappa_1 = a\}}.$$

        ii. *If $\zeta(t, x) = -1$, their type is set to be a if $\kappa_0 = \kappa_1 = a$ or $\kappa_0 \neq \kappa_1$ and A otherwise, so that for each $y \in B(x, r)$,*

$$w(t, y) = (1 - u)w(t_-, y) + u\left(\mathbf{1}_{\{\kappa_0 = \kappa_1 = a\}} + \mathbf{1}_{\{\kappa_0 \neq \kappa_1\}}\right).$$

Existence of the SLFVFS is guaranteed by the methods of *ETHERIDGE, VÉBER and YU* (2019, [15]). Indeed, we could have taken different measures μ and ν_r according to whether events are selective or neutral. Although it is convenient to take the strength of selection to be constant in space and have its direction determined by the variable $\zeta \in \{-1, +1\}$, we could, of course, have defined a much more general model. For example, one could allow $\mathbf{s}$ to vary in space, or even resample $\mathbf{s}\zeta$ from a suitable random field whenever the environment is resampled. However, this would be at the expense of considerably more complicated notation and it would become more involved to exploit the Poisson structure of our model. See Remark 4.4 below for some comments on when our scaling result would generalise.

One of the key tools in the study of the neutral SLFV is the dual process of coalescing random walkers which traces out the genealogical trees relating individuals in a sample from the population. An ancestral lineage doesn't move until it is both in the region affected by an event and is among the offspring of that event, at which time it jumps to the location of the parent of the event (which is uniformly distributed on the affected region). Things are more complicated in the presence of selection. Whereas in the neutral case we can always identify the distribution of the location of the parent of each event, now, at a selective event, even knowing the state of the environment, we are unable to identify which of the 'potential parents' is the true parent of the event without knowing their types. These can only be established by tracing further into the past. The resolution is to follow all potential ancestral lineages backwards in time. This results in a system of branching and coalescing walks in which branching and coalescence events are 'tagged' according to the state of the environment at the time at which they occur.

Just as in the neutral case, the dynamics of the dual are driven by the time reversals $\overleftarrow{\Pi}^{neu}$, $\overleftarrow{\Pi}^{sel}$, $\overleftarrow{\Pi}^{env}$ of the Poisson point processes of events that drove the forwards in time process, that is

$$\overleftarrow{\Pi}^i := \left\{ (-t, x, r, u) : (t, x, r, u) \in \Pi^i \right\}, \quad i \in \{\text{neu}, \text{sel}, \text{env}\}.$$

It is convenient to extend time to the whole of $\mathbb{R}$. The distribution of these Poisson point processes is then invariant under the time reversal.

We emphasize that time for the process of ancestral lineages runs in the opposite direction to that for the allele frequencies. Our dual will relate the distribution of allele frequencies in a sample from the population at a time T, to allele frequencies at time 0. More precisely, suppose that we know the frequencies $w(0, \cdot)$ of a-alleles at time 0. At time T, which we think of as 'the present', we sample j individuals from locations $\chi_0^1, \ldots, \chi_0^j$. Tracing backwards in time, we write $\chi_s^1, \ldots, \chi_s^{N_s}$ for the locations of the N_s potential ancestors that make up our dual at time s before the present.

Definition 4.2 (Ancestral selection graph). *We first define a $\bigcup_{n \geq 1} (\mathbb{R}^d)^n$-valued Markov process, $((\chi_t^i)_{i=1}^{N_t})_{t \geq 0}$, enriched by 'environmental tags' as follows:*
At each $\tau \in \overleftarrow{\Pi}^{env}$ the environment is resampled;
At each event $(t, x, r, u) \in \overleftarrow{\Pi}^{neu}$,

1. *for each $\chi_{t_-}^i \in B(x, r)$, independently mark the corresponding potential ancestor with probability u;*

2. *if at least one lineage is marked, all marked lineages disappear and are replaced by a single potential ancestor, whose location is drawn uniformly at random from within $B(x, r)$.*

At each event $(t, x, r, u) \in \overleftarrow{\Pi}^{sel}$:

1. *for each $\chi_{t_-}^i \in B(x, r)$, independently mark the corresponding potential ancestor with probability u;*

2. *if at least one lineage is marked, all marked lineages disappear and are replaced by* two *potential ancestors, whose locations are drawn independently and uniformly from within $B(x, r)$. The type of the environment is recorded.*

In both cases, if no particles are marked, then nothing happens.
To determine the distribution of types of a sample of the population $w(T, \cdot)$, taken from locations $\chi_0^1, \ldots, \chi_0^j$, knowing the distribution of $w(0, \cdot)$ at time T before the present, first evolve the process of branching and coalescing lineages until time T. At time T assign types to $\chi_T^1, \ldots, \chi_T^{N_T}$ using independent Bernoulli random variables such that $\mathbb{P}[\texttt{Type}(\chi_T^i) = a] = w(0, \chi_T^i)$. Tracing back through the system of branching and coalescing lineages $(\chi_t^i)_{i=1}^{N_t}$, we define types recursively: at each neutral event the lineages that coalesced during the event are assigned the type of the parent; at a selective event, if the state of the environment is equal to 1 then all coalescing lineages are type a if and only if both parents are type a, otherwise they are type A whereas if the state of the environment is equal to -1, all coalescing lineages are type A if and only if both parents are type A, otherwise they are type a. The distribution of types at time zero is the desired quantity.

Since we only consider finitely many initial individuals in the sample, the jump rate in this process is finite and so this description gives rise to a well-defined process.

This dual process is the analogue for the SLFVFS of the Ancestral Selection Graph (ASG), introduced in the companion papers *KRONE and NEUHAUSER* (1997, [27]), *NEUHAUSER and*

13

KRONE (1997, [37]), which describes all the potential ancestors of a sample from a population evolving according to the Wright-Fisher diffusion with selection. Indeed we could be more careful and use this process to extract the genealogy of a sample from the population. However, in this setting, this object seems to be rather unwieldy and, under the scalings in which we are interested, it will not converge to a well-defined limit.

Remark 4.3. *Informally, the procedure described above allows us to write down an expression, in terms of the marked process of branching and coalescing ancestral lineages and $w(0, \cdot)$, for $\mathbb{E}[\prod_{i=1}^{j} w(T, \chi_0^i)]$; that is the probability that j individuals, sampled from the present day population at locations $\chi_0^1, \ldots, \chi_0^j$, are all of type a. More formally, since the density of the SLFVFS is only defined Lebesgue-almost everywhere, the quantities $w(T, \chi_0^i)$ are only defined for Lebesgue almost every choice of $\chi_0^1, \ldots, \chi_0^j$ and, just as in ETHERIDGE, VÉBER and YU (2019, [15]) Section 1.2, this duality must be defined 'weakly', that is by integrating against a suitable test function $\psi(\chi_0^1, \ldots, \chi_0^j)$. Also mirroring that setting, the resulting 'moment duality' is sufficient to guarantee uniqueness of the SLFVFS. Since we do not use the duality in what follows, we refer the reader to ETHERIDGE, VÉBER and YU (2019, [15]) for details.*

4.2 Scaling the SLFVFS

We are interested in the effects of fluctuating selection over large spatial and temporal scales and so we shall consider a rescaling of our model. *ETHERIDGE, VÉBER and YU* (2019, [15]) consider the corresponding process in which selection does not fluctuate with time, but instead always favours type A (say). In that setting it is shown that if impact scales as $u_n = \bar{u}/n^{1/3}$ and selection scales as $\mathbf{s}_n = \mathbf{s}/n^{2/3}$, then as $n \to \infty$, $w(nt, n^{1/3}x)$ (or rather a local average of this quantity) converges in $d \geq 2$ to the solution to the deterministic Fisher-KPP equation, and to the solution of the corresponding stochastic p.d.e. in which a 'Wright-Fisher noise' term, corresponding to genetic drift, has been added in $d = 1$. Here we wish to consider short periods of stronger selection and so, by analogy with what we did in Section 3, we choose $\mathbf{s}_n = \mathbf{s}n^\alpha/n^{2/3}$ for some $\alpha > 0$, but we change the favoured type at times of mean $1/n^{2\alpha}$. This is of course the scaling suggested by the Central Limit Theorem (and is the natural analogue of our results in Section 3). In order to be able to ignore terms of order $n\mathbf{s}_n^2 u_n$ (as we did in the nonspatial setting) it is easy to see that we must take $\alpha \in (0, 1/3)$. In fact, our proof of our scaling result will require $\alpha \in (0, 1/6)$.

We must also scale the environment in a consistent way. In the examples that we have in mind, environmental correlations can be expected to extend over very large scales and so we actually fix the correlations in the *limiting* environment by fixing the distribution of a random field ξ and at the nth stage of the scaling sampling the environment according to ξ_n determined by

$$\xi(x) = \xi_n(n^{-1/3}x).$$

At the nth stage of the rescaling, the environment will be resampled at points of a rate $n^{2\alpha}$ Poisson process.

Remark 4.4 (Extensions)**.** *We have taken selection to be constant in magnitude and just to vary in sign. This is not necessary, even for our scaling result. As an obvious extension, we could fix the distribution of $\mathbf{s}(x)\xi(x)$ and, at the nth stage of the scaling define*

$$\mathbf{s}_n(n^{-1/3}x)\xi_n(n^{-1/3}x) = \frac{n^\alpha}{n^{2/3}}\mathbf{s}(x)\xi(x),$$

At the expense of introducing an additional truncation of $\mathbf{s}(x)$ at the nth stage of the scaling, to ensure that $|\mathbf{s}_n(x)\xi_n(x)| < 1$, it is enough to insist that

$$\mathbb{P}[\mathbf{s}(x)\xi(x) > z] = \mathbb{P}[\mathbf{s}(x)\xi(x) < -z], \quad \forall z \in \mathbb{R},$$

and

$$\mathbb{E}[|\mathbf{s}(x)\xi(x)|^{2+\epsilon}] < \infty, \qquad \text{for some } \epsilon > 0,$$

plus some regularity to reflect equation (4.3) below. To avoid a proliferation of notation, we omit this somewhat artificial generalisation of our results.

Our definition of the SLFVFS is still rather general. We include it to underline the possibility of extending our results. However, in the interest of avoiding even more complex expressions than those that follow, from now on we shall specialise to fix the radius and impact of reproduction events.

Assumption 4.5. *From now on, fix $R \in (0, \infty)$ and $\bar{u} \in (0, 1)$ and take*

$$\mu(dr) = \delta_R(dr), \qquad \nu_r(du) = \delta_{\bar{u}}(du).$$

Just as in *ETHERIDGE, VÉBER and YU* (2019, [15]), we shall prove convergence, not of the sequence of densities of the SLFVFS, but of a sequence of local averages. We require some notation. Let R be the fixed radius of events. Set

$$R_n = n^{-1/3}R, \quad B_n(x) = B(x, R_n),$$

and define the sequence of rescaled processes

$$w_n(t, x) = w(nt, n^{1/3}x), \quad \overline{w}_n(t, x) = \fint_{B_n(x)} w_n(t, y)dy, \tag{4.2}$$

where $\fint_{B_n(x)}$ denotes an average integral over the ball $B_n(x)$. We write V_R for the volume of a ball of radius R.

Theorem 4.6. *Write $(\overline{M}_t^n)_{t\geq 0}$ for the measure-valued process with density $(\overline{w}_n(t))_{t\geq 0}$. Suppose that $(M_0^n)_{n\geq 1}$ converges weakly in $\mathcal{M}_\lambda$ to the measure M_0^∞ with density $w^\infty(0, x) = \lim_{n\to\infty} \overline{w}_n(0, x)$. Further, fix $\alpha \in (0, 1/6)$, set $\mathbf{s}_n = \mathbf{s}n^\alpha/n^{2/3}$, $u_n = \bar{u}/n^{1/3}$ and suppose that the environment is resampled at the times of a Poisson process of rate $n^{2\alpha}$. We assume that the correlation function $g(x, y)$ that is determined by the environment satisfies*

$$|g(x, x) - g(x, y)| \leq C|x - y| \qquad \text{for all } x, y \in \mathbb{R}^d. \tag{4.3}$$

Then the sequence $(\overline{M}_{\cdot}^n)_{n\geq 1}$ is tight in $D([0, \infty), \mathcal{M}_\lambda)$ (the space of càdlàg functions on $[0, \infty)$ taking values in $\mathcal{M}_\lambda$). Moreover, for any weak limit point $(M_t^\infty)_{t\geq 0}$, writing w^∞ for a representative of the density of M^∞,

1. *in dimension $d = 1$, w^∞ is a process for which, for every $F \in C^\infty(\mathbb{R})$ and for every $f \in$*

$C_c^\infty(\mathbb{R})$,

$$F\left(\langle w^\infty(t), f\rangle\right) - F\left(\langle w^\infty(0), f\rangle\right)$$

$$- \int_0^t F'\left(\langle w^\infty(s), f\rangle\right)\left\{ \frac{\bar{u}\Gamma_R}{2}\langle w^\infty(s), \Delta f\rangle \right.$$

$$\left. + V_R^2\bar{u}^2\mathbf{s}^2\langle w^\infty(s)(1 - w^\infty(s))(1 - 2w^\infty(s)), f\rangle \right\} \mathrm{d}s$$

$$- \int_0^t F''\left(\langle w^\infty(s), f\rangle\right)$$

$$\times \left\{ \iint_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left[V_R^2\bar{u}^2\mathbf{s}^2 g(x,y)w^\infty(s,x)(1 - w^\infty(s,x)) \right. \right.$$

$$\left. \times w^\infty(s,y)(1 - w^\infty(s,y))f(x)f(y)\right]\mathrm{d}x\mathrm{d}y$$

$$\left. + \frac{\bar{u}^2 V_R^2}{2}\langle w^\infty(s)(1 - w^\infty(s)), f^2\rangle \right\}\mathrm{d}s \quad (4.4)$$

is a continuous martingale;

2. *in dimension $d \geq 2$, w^∞ is the process for which, for every $F \in C^\infty(\mathbb{R})$ and for every $f \in C_c^2\left(\mathbb{R}^d\right)$,*

$$F\left(\langle w^\infty(t), f\rangle\right) - F\left(\langle w^\infty(0), f\rangle\right)$$

$$- \int_0^t F'\left(\langle w^\infty(s), f\rangle\right)\left\{ \frac{\bar{u}\Gamma_R}{2}\langle w^\infty(s), \Delta f\rangle \right.$$

$$\left. + V_R^2\bar{u}^2\mathbf{s}^2\langle w^\infty(s)(1 - w^\infty(s))(1 - 2w^\infty(s)), f\rangle \right\}\mathrm{d}s$$

$$- \int_0^t F''\left(\langle w^\infty(s), f\rangle\right)$$

$$\times \left\{ \iint_{\mathbb{R}^d}\int_{\mathbb{R}^d}\left[V_R^2\bar{u}^2\mathbf{s}^2 g(x,y)w^\infty(s,x)(1 - w^\infty(s,x)) \right. \right.$$

$$\left. \times w^\infty(s,y)(1 - w^\infty(s,y))f(x)f(y)\right]\mathrm{d}x\mathrm{d}y \right\}\mathrm{d}s \quad (4.5)$$

is a continuous martingale. Moreover, the solution to this martingale problem is unique and so $\{\overline{M}^n\}_{n\geq 1}$ actually converges.

The constant Γ_R depends only on R and is defined in (8.19).

The proof of uniqueness in $d \geq 2$ uses a uniqueness result of *HU, NUALART and SON* [24] for a corresponding stochastic p.d.e.. In Appendix B, we follow the approach of *KURTZ* (2010, [31]), which uses the Markov Mapping Theorem, to show that any solution to the martingale problem (4.5) is actually a weak solution to the stochastic p.d.e.:

$$\mathrm{d}w^\infty = \left(\frac{\bar{u}\Gamma_R}{2}\Delta w^\infty + \bar{u}^2 V_R^2\mathbf{s}^2 w^\infty(1 - w^\infty)(1 - 2w^\infty) \right)\mathrm{d}t$$

$$+ \sqrt{2}\bar{u}V_R\mathbf{s}w^\infty(1 - w^\infty)W(\mathrm{d}t, \mathrm{d}x), \quad (4.6)$$

where the noise W is white in time and coloured in space, with quadratic variation given by

$$\langle W(\phi) \rangle_t = t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(x,y) \phi(x) \phi(y) \mathrm{d}x \mathrm{d}y. \tag{4.7}$$

The corresponding equation for dimension $d = 1$ is

$$dw^\infty = \left(\frac{\bar{u}\Gamma_R}{2} \Delta w^\infty + 4R^2 \bar{u}^2 \mathbf{s}^2 w^\infty (1 - w^\infty)(1 - 2w^\infty) \right) dt$$
$$+ \sqrt{2} 2R\bar{u}\mathbf{s} w^\infty (1 - w^\infty) W(\mathrm{d}t, \mathrm{d}x) + 2R\bar{u}\sqrt{w^\infty (1 - w^\infty)} \mathcal{W}(\mathrm{d}t, \mathrm{d}x), \tag{4.8}$$

where W is white in time and coloured in space as above and $\mathcal{W}$ is a space-time white noise. As in $d \geq 2$, any solution to the martingale problem (4.4) will be a weak solution to this stochastic p.d.e., but we only have a proof of uniqueness of (4.8) in the special case in which W is also space-time white noise (in which case we can invoke the duality of Section 5).

5 Duality

In general, we have been unable to identify a useful dual process for our limiting equation. The exceptions are the non-spatial setting and the special case of one spatial dimension, with both noises in the stochastic p.d.e. being white in space as well as time. In order to obtain these duals, we transform the system in a way inspired by *BLATH, ETHERIDGE and MEREDITH* (2007, [8]).

Consider first the non-spatial case. We rewrite equation (3.7) by making the substitution $X = 1 - 2p$. The process X, which takes values in $[-1, 1]$, satisfies

$$\mathrm{d}X_t = \frac{1}{2} \bar{u}^2 \mathbf{s}^2 \left(X_t^3 - X_t \right) \mathrm{d}t + \bar{u}\sqrt{1 - X_t^2} \mathrm{d}B_t^1 + \frac{\sqrt{2}}{2} \bar{u}\mathbf{s}(1 - X_t^2) \mathrm{d}B_t^2, \tag{5.1}$$

for independent Brownian motions B^1, B^2.

Lemma 5.1. *The solution to the transformed equation* (5.1) *is dual to a branching annihilating process* $\{N_t\}_{t \geq 0}$ *with transitions*

1. $N \mapsto N + 2$ *at rate*

$$\frac{\bar{u}^2 \mathbf{s}^2}{2} \left(N + \binom{N}{2} \right);$$

2. For $N \geq 2$, $N \mapsto N - 2$ *at rate*

$$\bar{u}^2 (1 + \frac{\mathbf{s}^2}{2}) \binom{N}{2}.$$

The duality relationship takes the form

$$\mathbb{E}_{X_0} \left[X_t^{N_0} \right] = \mathbb{E}_{N_0} \left[X_0^{N_t} \right]$$

where the expectation on the left is with respect to the law of $\{X_t\}_{t \geq 0}$ *started from initial condition* X_0, *and that on the right is with respect to the law of* $\{N_t\}_{t \geq 0}$, *started from* N_0.

The proof is an application of Itô's formula. It is easy to see that started from an even number of particles, the dual process will die out in finite time (count the number of pairs of particles and compare to a subcritical birth-death process), corresponding to the process $\{p_t\}_{t\geq 0}$ of allele frequencies being absorbed in either zero or one. Of course, this is also readily checked directly for the diffusion (1.1) using the theory of speed and scale, but if we could find an analogous dual for a spatially extended population, where the theory of one dimensional diffusions is no longer helpful, we might be able to exploit it to study the behaviour of allele frequencies.

In one spatial dimension, if the noises W and $\mathcal{W}$ are both white in space as well as time, then we can extend this.

Lemma 5.2. *Suppose that $d = 1$ and w^∞ solves*

$$\mathrm{d}w^\infty = \left(\frac{\bar{u}\Gamma_R}{2}\Delta w^\infty + 4R^2\bar{u}^2\mathbf{s}^2 w^\infty(1-w^\infty)(1-2w^\infty)\right)\mathrm{d}t$$
$$+ \sqrt{2}2R\bar{u}\mathbf{s}w^\infty(1-w^\infty)W(\mathrm{d}t,\mathrm{d}x) + 2R\bar{u}\sqrt{w^\infty(1-w^\infty)}\mathcal{W}(\mathrm{d}t,\mathrm{d}x),$$

where W and $\mathcal{W}$ are independent space-time white noises. Then setting $X_t(x) = 1 - 2w^\infty(t,x)$, $\{X_t(x)\}_{t\geq 0}$ is dual to a system of branching-annihilating Brownian particles whose spatial locations at time t we denote by $\chi_1(t),\ldots,\chi_{N_t}(t)$, and whose dynamics are described as follows:

1. *Each particle, independently, follows a Brownian motion in $\mathbb{R}$, with diffusion constant $\bar{u}\Gamma_R$;*

2. *Each particle, independently, splits into three at rate $\bar{u}^2\mathbf{s}^2/2$;*

3. *Each pair of particles annihilates, at a rate $\bar{u}^2(1+\mathbf{s}^2/2)$, measured by their intersection local time;*

4. *Each pair of particles replicates (i.e. is replaced by two identical pairs) at rate $\bar{u}^2\mathbf{s}^2)/2$, also measured by their intersection local time.*

The duality is expressed for points $\chi_1(0),\ldots\chi_{N_0}(0) \in \mathbb{R}$ and any continuous function $w^\infty(0,\cdot) : \mathbb{R} \to [0,1]$, through

$$\mathbb{E}\left[\prod_{i=1}^{N_0} w_t^\infty(\chi_i(0))\right] = \mathbb{E}\left[\prod_{i=1}^{N_t} w_0^\infty(\chi_i(t))\right], \tag{5.2}$$

where the expectation on the left is with respect to the law of the stochastic p.d.e. and that on the right with respect to the law of the dual system of branching and annihilating lineages.

TRIBE (1995, [47]) gives a construction of the analogous system of coalescing Brownian motions, which is dual to the stochastic heat equation with Wright-Fisher noise, as discussed in *SHIGA* (1988 [45]). *DOERING, MUELLER* and *SMEREKA* (2002, [11]) provide a complete derivation in that context; see also *LIANG* (2009, [33]). We also note that *BIRKNER* (2003, [7]) considers a similar system of branching random walkers on $\mathbb{Z}^d$, in which particles reproduce at a rate that depends on the number of other particles within the same site.

Remark 5.3. *If we consider subdivided populations (i.e. an analogous model on a lattice), then the analogous system of stochastic (ordinary) differential equations satisfies a duality of this form with a system of branching and annihilating random walks. Moreover, we can apply the analogous transformation to the stochastic p.d.e. with coloured noise, but now, when we seek a branching*

annihilating dual process, in addition to the branching annihilating term when dual particles meet, we obtain cross terms of the form

$$g(x,y)X_t(x)X_t(y)\big(1 - X_t(x)^2\big)\big(1 - X_t(y)^2\big).$$

We can rearrange the factors involving X as

$$\big(X_t(x)^2 X_t(y)^2 - X_t(x)X_t(y)\big) + (1 - X_t(x)X_t(y))$$
$$- \big(X_t(x)^2 - X_t(x)X_t(y)\big) - \big(X_t(y)^2 - X_t(x)X_t(y)\big). \quad (5.3)$$

If $g(x,y) > 0$, then we can interpret the first two terms in (5.3) as branching and annihilating terms in a putative dual; if $g(x,y) < 0$, then the last two terms can be interpreted as one particle jumping to the location of another. However, we have not, in either case, found a way to interpret both terms simultaneously. The obvious approach is to follow ATHREYA and TRIBE (2000, [1]) and introduce an additional 'marker' that switches sign every time we have an event of 'the wrong sign'. This leads to a Feynman-Kac correction term in the duality relation (5.2), which it turns out is infinite.

6 Tracer dynamics

On passing to a stochastic p.d.e. limit, we have lost sight of the way in which individuals in the population were related to one another and the ancestral selection graphs which encoded that information in the prelimiting models do not converge. However, some information about heredity can be recovered using the notion of 'tracers'. The idea, which finds its roots in the statistical physics literature, has more recently found application in models of population genetics, notably in *HALLATSCHEK and NELSON* (2008, [23]), or, for a more mathematical approach, see *DURRETT and FAN* (2016, [13]). The idea is simple: one labels some portion of the population of type a individuals, say, at time zero not just according to their type at the selected locus, but also with a 'neutral marker' that is passed down from parent to offspring. Individuals in the population at time t that carry the neutral marker are precisely the descendants of our original marked individuals.

To introduce this in our setting, let us label a portion of the a population at time zero by a neutral marker. We shall use $v(t,x)$ to denote the proportion of the (total) population that are both type a and labelled, and we shall use a^* to denote that combined type. Thus

$$\{\text{type } a^* \text{ individuals}\} \subseteq \{\text{type } a \text{ individuals}\}$$

and $v(t,x) \le w(t,x)$.

The dynamics are driven by the same Poisson point processes of events as before, but now we modify our description of inheritance to include the extra label.

Definition 6.1 (The SLFVFS with tracers). *Let ζ, Π^{neu}, Π^{fsel}, Π^{env} be exactly as in Definition 4.1. The dynamics of the pair (v,w) can be described as follows. Write $a^\dagger$ to denote individuals of type a, but not a^*.*

1. *If $(t,x,r,\bar{u}) \in \Pi^{neu}$, a neutral event occurs at time t within the closed ball $B(x,r)$. Then:*

   (a) *Choose a parental location l according to the uniform distribution over $B(x,r)$.*

   (b) *Choose the parental type $\kappa \in \{a^*, a^\dagger, A\}$ according to distribution*

   $$\mathbb{P}\left[\kappa = a^*\right] = v(t_-, l), \quad \mathbb{P}\left[\kappa = a^\dagger\right] = w(t_-, l) - v(t_-, l),$$
   $$\mathbb{P}\left[\kappa = A\right] = 1 - w(t_-, l).$$

(c) For each $y \in B(x, r)$,

$$v(t, y) = (1 - \bar{u})v(t_-, y) + \bar{u}\mathbf{1}_{\{\kappa = a^*\}}(y);$$
$$w(t, y) = (1 - \bar{u})w(t_-, y) + \bar{u}\mathbf{1}_{\{\kappa \in \{a^*, a^\dagger\}\}}(y).$$

2. *If $(t, x, r, \bar{u}) \in \Pi^{fsel}$, a selective event occurs at time t within the closed ball $B(x, r)$. Then:*

   (a) *Choose the two parental locations l_0, l_1 independently, according to the uniform distribution on $B(x, r)$.*

   (b) *Choose the two parental types, κ_0, κ_1, according to*

   $$\mathbb{P}\left[\kappa_i = a^*\right] = v(t_-, l_i), \quad \mathbb{P}\left[\kappa_i = a^\dagger\right] = w(t_-, l_i) - v(t_-, l_i),$$
   $$\mathbb{P}\left[\kappa_i = A\right] = 1 - w(t_-, l_i).$$

   (c) i. *If $\zeta(t, x) = 1$, offspring inherit type κ_0 if κ_0 and κ_1 are both type a (with or without the neutral marker), otherwise they are type A; so*

   $$v(t, y) = (1 - \bar{u})v(t_-, y) + \bar{u}\mathbf{1}_{\{\kappa_0 = a^*, \kappa_1 \in \{a^*, a^\dagger\}\}}$$
   $$w(t, y) = (1 - \bar{u})w(t_-, y) + \bar{u}\mathbf{1}_{\{\kappa_0, \kappa_1 \in \{a^*, a^\dagger\}\}}.$$

   ii. *If $\zeta(t, x) = -1$, offspring inherit type κ_0 if κ_0 is type a, and they inherit type κ_1 if κ_0 is type A. Thus*

   $$v(t, y) = (1 - \bar{u})v(t_-, y) + \bar{u}\left(\mathbf{1}_{\{\kappa_0 = a^*\}} + \mathbf{1}_{\{\kappa_0 = A, \kappa_1 = a^*\}}\right);$$
   $$w(t, y) = (1 - \bar{u})w(t_-, y) + \bar{u}\mathbf{1}_{\{\kappa_0 = \kappa_1 = A\}^c}.$$

Theorem 6.2. *Applying our previous scalings, let the population evolve under the assumptions of Theorem 4.6.*

Assume that the sequence of measures of the initial states of the marked population converges weakly in $\mathcal{M}_\lambda$ to a measure with a density $v^\infty(0, x) = \lim_{n \to \infty} \bar{v}_n(0, x)$. Then as $n \to \infty$, the corresponding sequence of rescaled measure-valued processes is tight in $D([0, \infty), \mathcal{M}_\lambda^2)$, and any limit point is a weak solution to a system of stochastic p.d.e.'s which in $d \geq 2$ takes the form

$$dw^\infty = \left(\frac{\bar{u}\Gamma_R}{2}\Delta w^\infty + V_R^2 \bar{u}^2 \mathbf{s}^2 w^\infty(1 - w^\infty)(1 - 2w^\infty)\right)dt + \sqrt{2}V_R \bar{u}\mathbf{s}w^\infty(1 - w^\infty)W(dt, dx),$$

$$dv^\infty = \left(\frac{\bar{u}\Gamma_R}{2}\Delta v^\infty + V_R^2 \bar{u}^2 \mathbf{s}^2 v^\infty(1 - w^\infty)(1 - 2w^\infty)\right)dt + \sqrt{2}V_R \bar{u}\mathbf{s}v^\infty(1 - w^\infty)W(dt, dx),$$

where the noise W is as before.

For dimension $d = 1$, the limiting process is a weak solution to the system of stochastic partial differential equations

$$dw^\infty = \left(\frac{\bar{u}\Gamma_R}{2}\Delta w^\infty + 4R^2 \bar{u}^2 \mathbf{s}^2 w^\infty(1 - w^\infty)(1 - 2w^\infty)\right)dt + \sqrt{2}2R\bar{u}w^\infty(1 - w^\infty)W(dt, dx)$$
$$+ 2R\bar{u}\sqrt{v^\infty(1 - w^\infty)}\mathcal{W}^0(dt, dx) + 2R\bar{u}\sqrt{(w - v)^\infty(1 - w^\infty)}\mathcal{W}^1(dt, dx),$$

$$dv^\infty = \left(\frac{\bar{u}\Gamma_R}{2}\Delta v^\infty + 4R^2 \bar{u}^2 \mathbf{s}^2 v^\infty(1 - w^\infty)(1 - 2w^\infty)\right)dt + \sqrt{2}2R\bar{u}\mathbf{s}v^\infty(1 - w^\infty)W(dt, dx)$$
$$+ 2R\bar{u}\sqrt{v^\infty(1 - w^\infty)}\mathcal{W}^0(dt, dx) + 2R\bar{u}\sqrt{v^\infty(w - v^\infty)}\mathcal{W}^2(dt, dx),$$

where, as before, W is white in time and coloured in space and and $\mathcal{W}^i$, $i = 0, 1, 2$, are independent space-time white noises.

The proof of this result is an even longer version of the proof of Theorem 4.6, but it follows exactly the same strategy. First we show that the limit points are solutions to appropriate martingale problems, then we show that any solution to the martingale problem provides a weak solution to the system of stochastic p.d.e.'s. Rather than providing details of the proof, we indicate why this result is to be expected. We once again use the trick of *KURTZ* (1973, [28]). In the case of one dimension and genic selection, *DURRETT and FAN* (2016, [13]) obtain a pair of stochastic p.d.e.'s of the form

$$
\begin{aligned}
\mathrm{d}w &= (\alpha\Delta w + \mathbf{s}w(1-w))\,\mathrm{d}t + \sqrt{v(1-w)}W^0(\mathrm{d}t,\mathrm{d}x) + \sqrt{(w-v)(1-w)}W^1(\mathrm{d}t,\mathrm{d}x) \\
\mathrm{d}v &= (\alpha\Delta v + \mathbf{s}v(1-w))\,\mathrm{d}t + \sqrt{v(1-w)}W^0(\mathrm{d}t,\mathrm{d}x) + \sqrt{v(w-v)}W^2(\mathrm{d}t,\mathrm{d}x),
\end{aligned}
$$

where W^0, W^1 and W^2 are independent space-time white noises. Writing the corresponding generator acting on test functions $f(w,l)$ as $\mathcal{L}^{neu} + \mathcal{L}^{sel}$ as in Section 3, to identify the generator in the limit of (appropriately scaled) rapidly fluctuating selection we must evaluate $\mathcal{L}^{sel}(\mathcal{L}^{sel}f(\cdot,\cdot))(w,v)$ which, up to constants, is

$$
\left(w(1-w)\frac{\partial}{\partial w} + v(1-w)\frac{\partial}{\partial v}\right)\left\{w(1-w)\frac{\partial f}{\partial w} + v(1-w)\frac{\partial f}{\partial v}\right\}
$$
$$
= w(1-w)(1-2w)\frac{\partial f}{\partial w} + v(1-w)(1-2w)\frac{\partial f}{\partial v}
$$
$$
+ w^2(1-w)^2\frac{\partial^2 f}{\partial w^2} + 2wv(1-w)^2\frac{\partial^2 f}{\partial v\partial w} + v^2(1-w)^2\frac{\partial^2 f}{\partial v^2}.
$$

From this we see that the stochastic p.d.e's in Theorem 6.2 are of precisely the form that we should expect.

7 Numerical results

In order to gain a little more intuition about the effects of fluctuating selection on allele frequencies, in this section we present the results of a simple numerical experiment. It is certainly not an exhaustive study, but it points to some of the challenges that face us in distinguishing causes of patterns of allele frequencies. Our simulations are not of the spatial Lambda-Fleming-Viot models, but of natural extensions of the classical Moran model to incorporate spatial structure and fluctuating selection. After suitable scaling, we expect allele frequencies under these models to converge to the same limiting stochastic p.d.e. as our scaled SLFVFS.

Definition 7.1 (Spatial Moran model with fluctuating selection)**.** *The population, which consists of two genetic types, $\{a, A\}$, lives at the vertices of a discrete lattice $\mathbb{L}$. There are N_d individuals at each vertex (or* deme*). The state of the environment at time t in deme x is denoted by $\zeta(t,x)$.*
  The dynamics of the process are described as follows:

1. **Reproduction events**

   (a) **Neutral events:** *For each deme, independently, at rate $\binom{N_d}{2}$ a pair of individuals is chosen (uniformly at random), one of the pair (picked at random) dies and the other splits in two;*

   (b) **Selective events:** *For each deme x, independently, at rate $\mathbf{s}N_d$ a pair of individuals is chosen (uniformly at random), one individual splits in two and the other one dies; if $\zeta(t,x) = -1$ and at least one of the pair is type a, then it is a type a individual that is chosen to split, whereas if $\zeta(t,x) = 1$ and at least one of the pair is type A, then a type A individual is chosen to split.*

2. **Migration events:** *For each pair of demes x_1, x_2, we associate a nonnegative parameter m_{x_1,x_2}. Independently for each pair, at rate $m_{x_1,x_2} N_d$, an individual is chosen uniformly at random from each of the demes x_1, x_2 and they exchange places.*

3. **Environmental events:** *At the times of a Poisson process of rate α, which is independent of those driving reproduction and migration, the environment is resampled. The value of the environment variable at each deme is uniformly distributed on $\{-1, +1\}$ and $\mathbb{E}[\zeta(t,x)\zeta(t,y)] = g(x,y)$ for a correlation function g.*

In the experiments that follow, we take the lattice $\mathbb{L}$ to be a circle of 100 demes with nearest neighbour migration at rate 1. We set $N_d = 400$, $\mathbf{s} = 0.1$ and $\alpha = 10$. The environment variables in demes $0 - 50$ all take the same value, as do those in demes $51 - 100$. We consider four different scenarios:

1. Demes $0 - 50$ and $51 - 100$ are perfectly anticorrelated; the environment fluctuates in time and the direction of selection is always different in demes $0 - 50$ and $51 - 100$;

2. Demes $0 - 100$ are perfectly correlated; the direction of selection fluctuates in time but it is the same in every deme;

3. Constant selection (the environment is fixed), with the direction of selection in demes $0 - 50$ the opposite of that in demes $51 - 100$;

4. The neutral case.

To ensure comparability of experiments, the same events are used for all the scenarios, with the only difference lying in the value of the environment variable. Thus, in the neutral case, either individual is equally likely to be the parent in 'selective events', irrespective of type. A more precise description of the code used for simulations can be found in Appendix D.

As a first comparison, Figure 1 shows the proportion of type a individuals across the whole population. This is just a single realisation of the experiment. There is certainly no dramatic divergence from neutrality. However, as we illustrate below, this can mask some more interesting effects at the local level.

In Figure 2, we have used a greyscale to record the proportion of type a individuals in each deme - the darker the colour, the greater the proportion of type a. In the top left, selection is fixed, and we clearly see the effect of type a being favoured in demes $50 - 100$. In the next two frames (top right and bottom left), the environment fluctuates, but whereas on the top right demes $0 - 50$ always favour the opposite type to demes $51 - 100$, on the bottom left all demes always favour the same type. The neutral model is the bottom right. When we repeat over many realisations, we see a greater concentration of types than for the neutral model, but it is certainly not easy to distinguish between the two frames.

Just following the overall proportion of types in a deme is throwing away a lot of information, which may be available in genetic data, about the distribution of families. To explore this we used 'tracers', further investigated in the context of Spatial Lambda-Fleming-Viot processes in Section 6. In Figure 3 we mark individuals descended from the population in particular demes at time zero. The darker the colour, the higher the proportion of marked individuals. In a constant environment, see the left panel, the surviving family is well adapted to the environment in demes $50 - 100$, but has difficulty invading demes $1 - 50$, where it is not favoured. It also struggles to expand beyond deme 80. This turns out to be because of competition with the equally well adapted family that is descended from individuals that were in deme 84 at time zero (see the right hand panel).
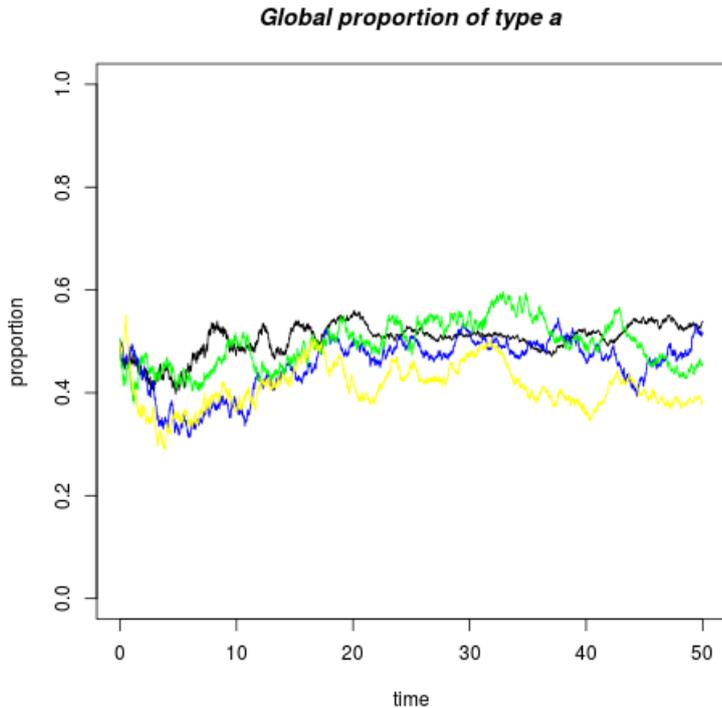
Figure 1: Global proportions of type a individuals; one realisation, four different scenarios: blue anticorrelated environments, green correlated environments, black constant selection, and yellow neutral. See the main text for a full explanation.

In the left panel of Figure 4, we see a successful family in a fluctuating environment (with demes $0 - 50$ and $51 - 100$ perfectly anticorrelated). The family is on the brink of extinction several times and is rescued by a change in the environment. In the right panel, we see a family that begins life right on the boundary between the two regions. The environments are perfectly anticorrelated and the family is able to survive and spread much more readily.

The right hand panel of Figure 4 can be contrasted with the left hand panel of Figure 5. The descendants of ancestors in deme 1 find it harder to spread in a perfectly correlated environment than in the perfectly anticorrelated environment of the previous figure. The trace of descendants of ancestors in deme 78 in the right hand panel of Figure 5 shows the 'thinning' of the family resulting from the periods of time when it is not favoured.

Finally, Figure 6 shows the trace of descendants of ancestors in demes 12 and 16 for the neutral model. There appears to be a barrier between the two families, which could easily be mistaken for a change in the environment somewhere between demes 12 and 16, on the lower side of which the family descended from deme 12 is better adapted and on the upper side of which the family from deme 16 is better adapted. In fact this is due to competition between two equally fit families.

8   Proof of Theorem 4.6

The proof of Theorem 4.6 will rest on Theorem 2.1 of *KURTZ* (1992, [29]) (or rather his Example 2.2). For a metric space E, let $l_m(E)$ be the space of measures on $[0, \infty) \times E$ such that $\mu \in l_m(E)$ if $\mu([0, t) \times E) = t$.
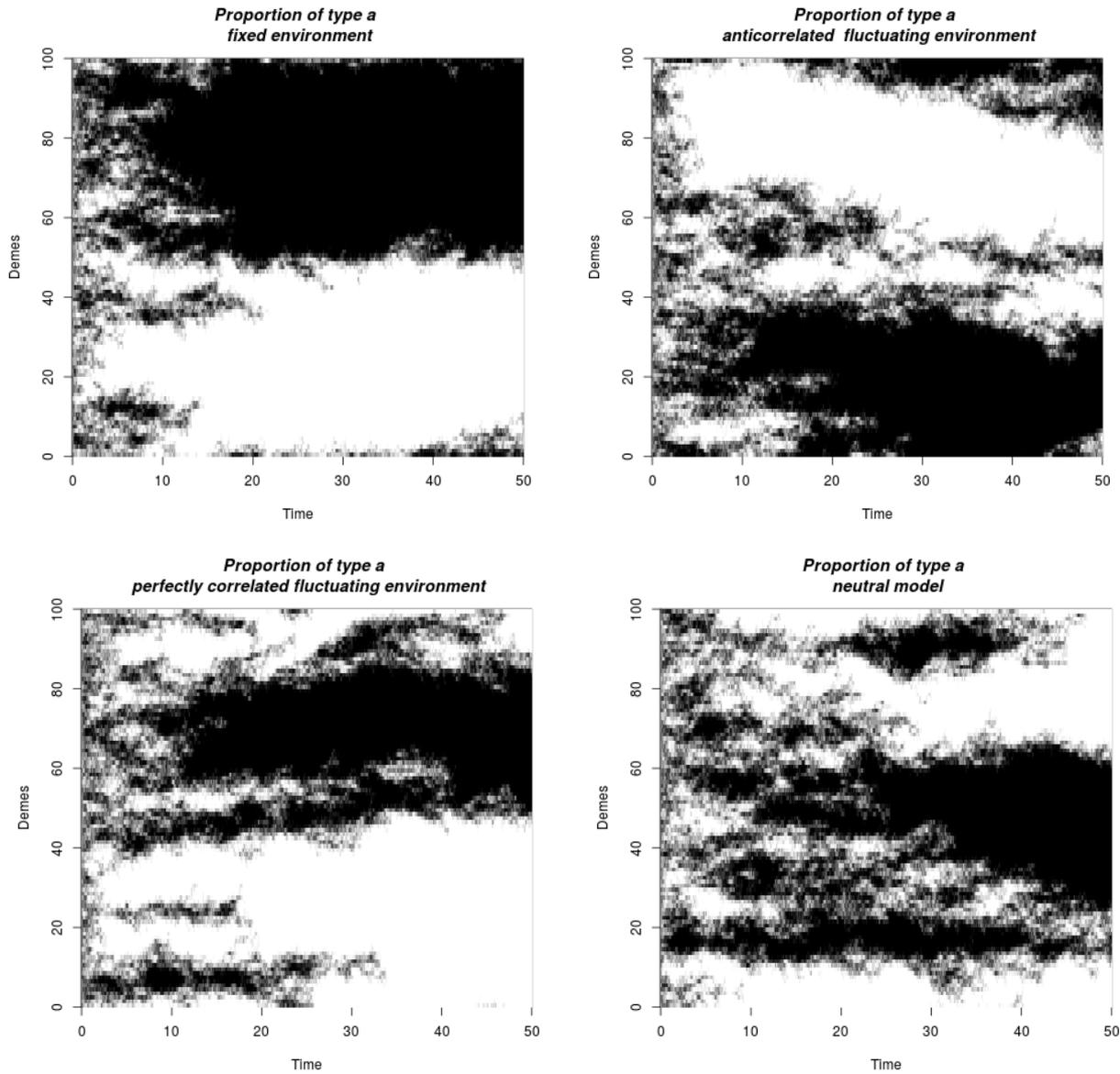
Figure 2: Map of the proportion of type a individuals in each deme; the darker the colour the higher the proportion of type a. Top left, constant selection; top right, fluctuating selection with demes $0-50$ and $51-100$ perfectly anticorrelated; bottom left, fluctuating selection with all demes perfectly correlated; bottom right, the neutral case.

24

Figure 3: Tracer map of individuals descended from ancestors located in particular demes at time zero; the darker the colour, the higher the proportion of such individuals. Fixed environment.

Figure 4: Tracer map of individuals descended from ancestors located in particular demes at time zero; the darker the colour, the higher the proportion of such individuals. The environment is fluctuating, with its value in demes $0 - 50$ perfectly anticorrelated with that in demes $51 - 100$.
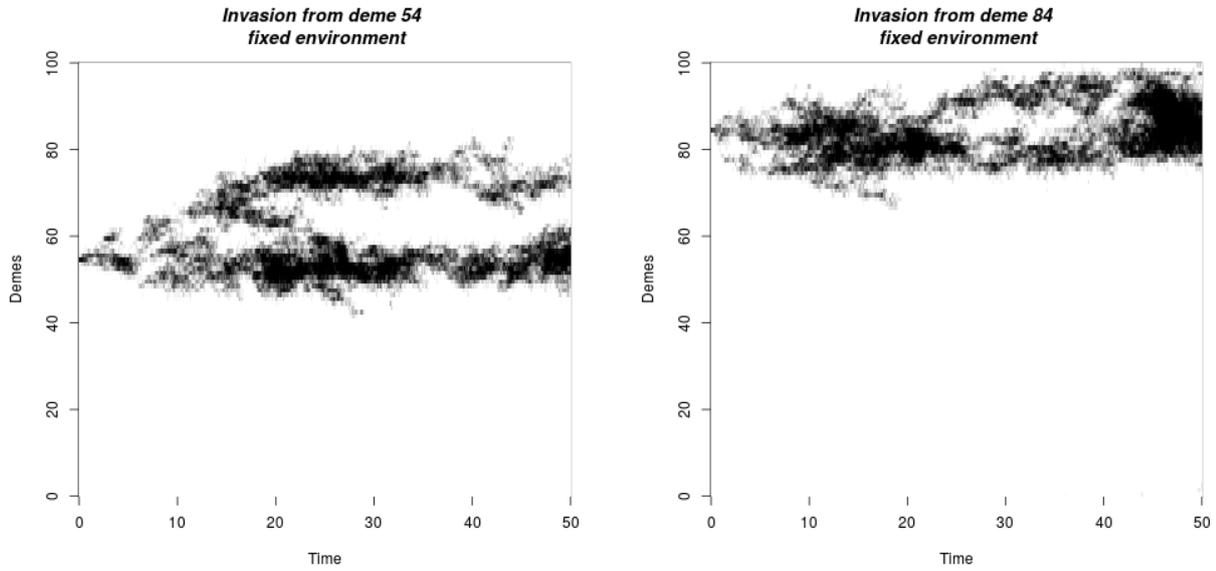
25

Figure 5: Tracer map of individuals descended from ancestors located in particular demes at time zero; the darker the colour, the higher the proportion of such individuals. The environment is fluctuating, with its value in demes $0 - 50$ perfectly correlated with that in demes $51 - 100$.
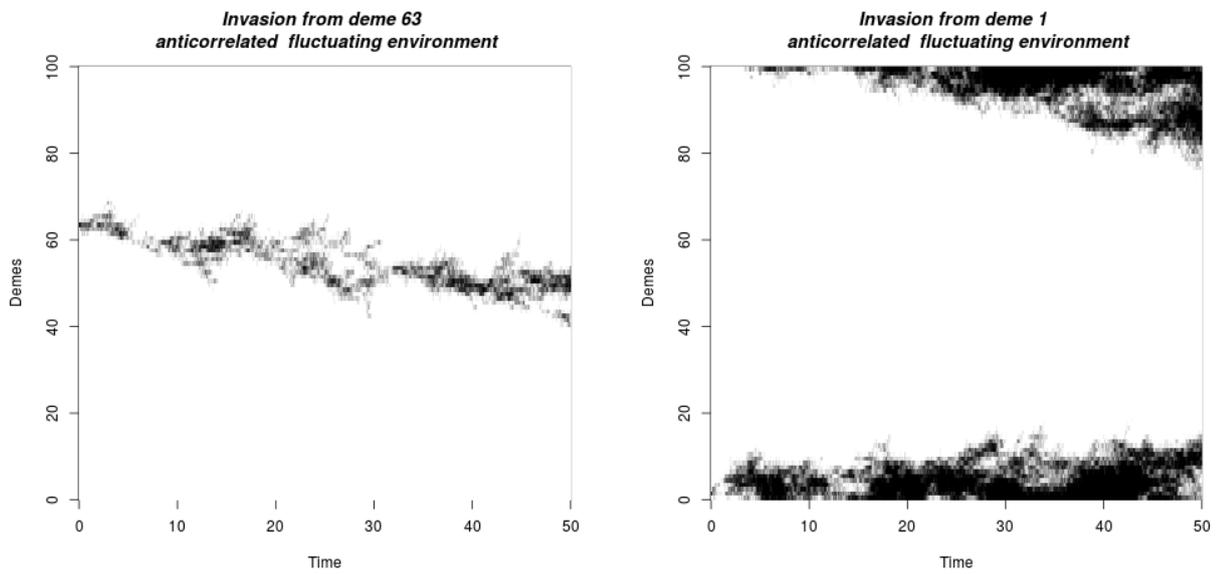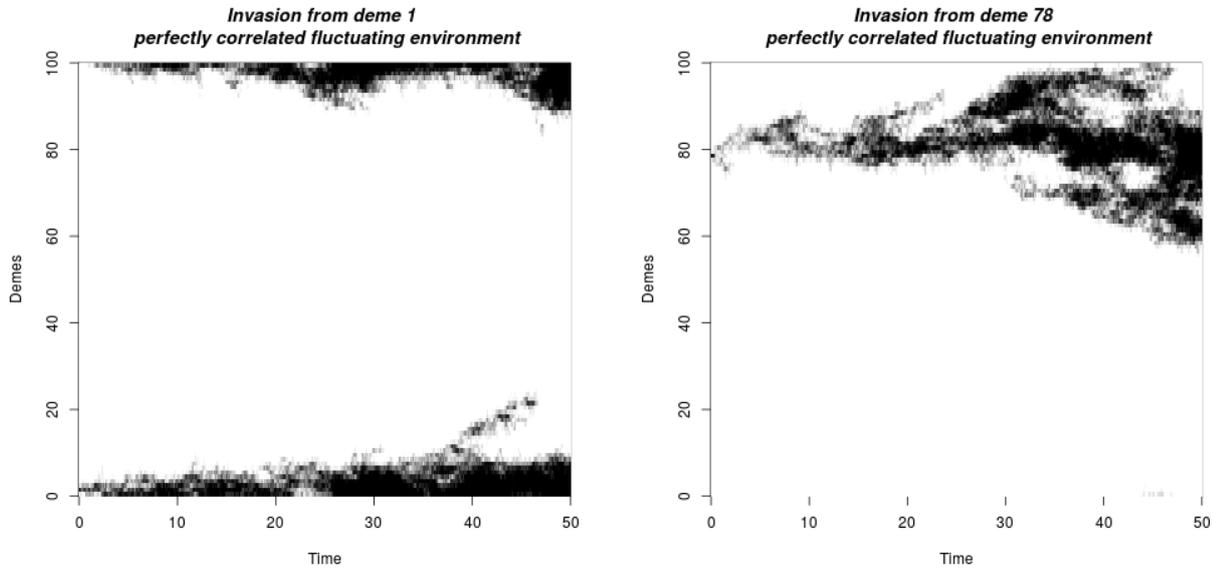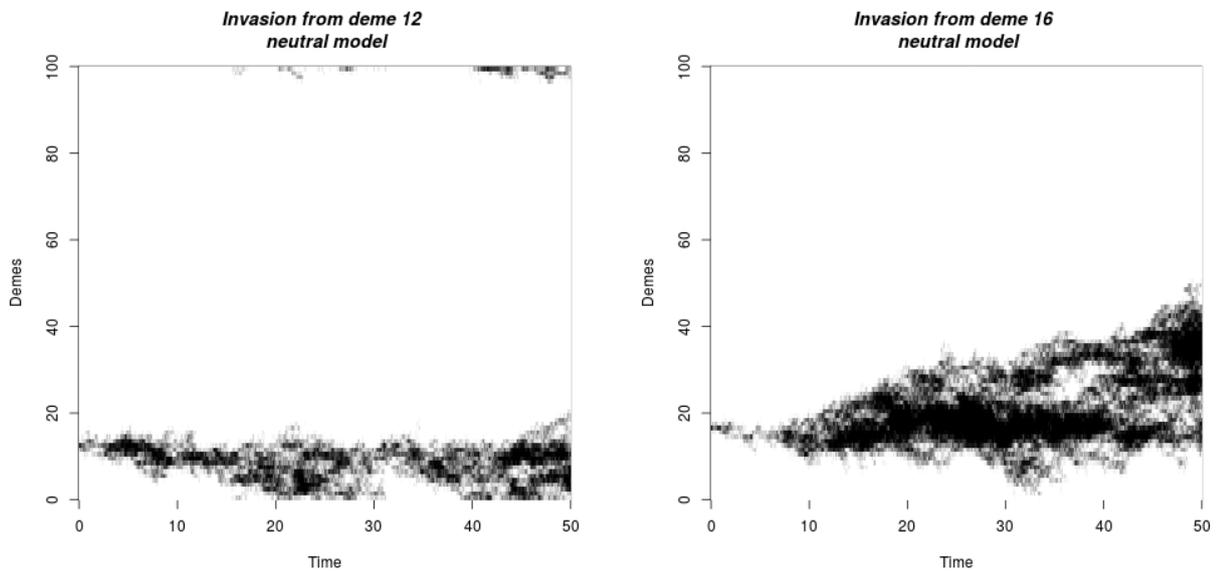

Figure 6: Tracer map of individuals descended from ancestors located in particular demes at time zero; the darker the colour, the higher the proportion of such individuals. Neutral model.

26

Theorem 8.1 (*KURTZ* (1992, [29]), *Theorem 2.1*)**.** *Let E_1, E_2 be complete separable metric spaces, and set $E = E_1 \times E_2$. For each n, let $\{(X_n, Y_n)\}$ be a stochastic process with sample paths in $D_E([0, \infty))$ adapted to a filtration $\{\mathcal{F}_t^n\}$. Assume that $\{X_n\}$ satisfies the compact containment condition, that is, for each $\epsilon > 0$ and $T > 0$, there exists a compact $K \subset E_1$ such that*

$$\inf_n \mathbb{P}[X_n(t) \in K, t \leq T] \geq 1 - \epsilon, \tag{8.1}$$

and assume that $\{Y_n(t) : t \geq 0, n = 1, 2, \ldots\}$ is relatively compact (as a collection of E_2-valued random variables). Suppose that there is an operator $A : \mathcal{D}(A) \subset \overline{C}(E_1) \to C(E_1 \times E_2)$ such that for $f \in \mathcal{D}(A)$ there is a process ϵ_n^f for which

$$f(X_n(t)) - \int_0^t Af(X_n(s), Y_n(s))\mathrm{d}s + \epsilon_n^f(t) \tag{8.2}$$

is an $\{\mathcal{F}_t^n\}$-martingale. Let $\mathcal{D}(A)$ be dense in $\overline{C}(E_1)$ in the topology of uniform convergence on compact sets. Suppose that for each $f \in \mathcal{D}(A)$ and each $T > 0$, there exists $p > 1$ such that

$$\sup_n \mathbb{E}\left[\int_0^T |Af(X_n(t), Y_n(t))|^p \mathrm{d}t\right] < \infty \tag{8.3}$$

and

$$\lim_{n \to \infty} \mathbb{E}\left[\sup_{t \leq T} |\epsilon_n^f(t)|\right] = 0. \tag{8.4}$$

Let Γ_n be the $l_m(E_2)$-valued random variable given by

$$\Gamma_n([0, t] \times B) = \int_0^t \mathbf{1}_B(Y_n(s))\mathrm{d}s.$$

Then $\{(X_n, \Gamma_n)\}$ is relatively compact in $D_{E_1}[0, \infty) \times l_m(E_2)$, and for any limit point (X, Γ) there exists a filtration $\{\mathcal{G}_t\}$ such that

$$f(X(t)) - \int_0^t \int_{E_2} Af(X(s), y)\Gamma(\mathrm{d}s \times \mathrm{d}y) \tag{8.5}$$

is a $\{\mathcal{G}_t\}$-martingale for each $f \in \mathcal{D}(A)$.

As a particular case, *KURTZ* (1992, [29]) provides the following example.

Example 8.2 (*KURTZ* (1992, [29]), *Example 2.2*)**.** *Suppose Y is stationary and ergodic, and Y_n in Theorem 8.1 is given by $Y_n(t) \equiv Y(nt)$. Then $\Gamma = m \times \pi$ where m denotes Lebesgue measure and π is the marginal distribution for Y. Consequently, under the assumptions of Theorem 8.1, X is a solution of the martingale problem for C given by*

$$Cf(x) = \int Af(x, y)\pi(\mathrm{d}y), \qquad f \in \mathcal{D}(A).$$

Let us briefly discuss how our setup fits into that of Theorem 8.1. The sequence of measure-valued evolutions $\left(\overline{M}_t^n\right)_{t \geq 0}$ corresponds to the process X_n, while the sequence of environments corresponds to the process Y_n. The sequence X_n takes values in $\mathcal{M}_\lambda$. The state space of the sequence Y_n can be identified with the product of $\{-1, 1\}$ with the collection of closed subsets of $\mathbb{R}^d$. Taking $\{E_k\}_{k \geq 1}$ to be an increasing sequence of closed and bounded subsets with $\mathbb{R}^d = \bigcup_{k=1}^\infty E_k$,

and writing H for Hausdorff distance, we set $H^\infty(A, B) = \sum_{k=1}^{\infty} \frac{1}{2^k} H(A \cap E_k, B \cap E_k)$, to recover a metric under which this latter space is complete and separable. Notice that Y_n defined in this way satisfies the assumptions of Example 8.2 with n replaced by $n^{2\alpha}$, and therefore those of Theorem 8.1. The domain of the generator A is given by the closure of

$$\mathcal{D}(A) = \left\{ F(\langle \cdot, f \rangle) : f \in C_c^\infty(\mathbb{R}^d), F \in C^\infty(\mathbb{R}) \right\}. \tag{8.6}$$

Since the state space of $\left(\overline{M}_t^n\right)_{t>0}$ is compact, the compact containment condition (8.1) is satisfied.

The key step in applying Theorem 8.1 is the identification of the generator A, which leads to the decomposition (8.2). This consists of two steps. First we approximate the part of the SLFVFS generator corresponding to neutral events, and second we deal with the terms corresponding to fluctuating selection. These two steps are accomplished in Subsection 8.1.

Our approach mirrors the strategy of Section 3. Having transformed the generator of the SLFVFS into a form which allows us to consider the process of averaged densities $\overline{w}_t^n$, we use the 'separation of timescales' trick of *KURTZ* (1973, [28]), to (up to a small error) split the generator into parts corresponding to neutral and selective events in a convenient way; see equation (8.16). We then consider the two terms in this decomposition. The analysis of the neutral part of the generator will closely follow *ETHERIDGE, VÉBER and YU* (2019, [15]); the novel step is the second, which deals with the terms corresponding to fluctuating selection. Our analysis will not only identify the correct form of A and ϵ_n^f (see (8.28) and (8.29) respectively), but also provides a uniform bound on A, which shows that (8.3) is satisfied.

Having identified A, and ϵ_n^f, we show that (8.4) holds. This bound relies on the Lipschitz bound (4.3) on the correlation kernel $g(x, y)$. Subsection 8.1 concludes with a discussion of the continuity estimates on the expectations of averaged densities of the measure-valued evolutions that are required to deduce our result. These estimates are stated as Equation (8.30) and Lemma 8.4 (and proved in Appendix A). This leads to a complete characterisation of limits of the sequence of averaged processes $\overline{w}_n$ as solutions to a martingale problem.

In Subsection 8.2 we discuss the uniqueness of solutions to the limiting martingale problem, which we have only managed to prove in dimensions $d \geq 2$. This requires a separate argument based on the fact that, under certain regularity conditions (which are satisfied as a result of Lemma 8.4), any solution to the martingale problem is a weak solution to a corresponding stochastic p.d.e. The weak uniqueness of solutions to the stochastic p.d.e. is a consequence of a uniqueness result of *HU, NUALART and SON* [24].

8.1 Identifying the limit

In what follows, C is a constant that depends on certain fixed quantities such as the functions f and F that appear in our test functions, c.f. (8.6), and the quantities $\mathbf{s}$ and $\bar{u}$ appearing in our scaled selection and impact parameters in the statement of Theorem 4.6. The quantity C may vary from line to line; its exact value is unimportant.

Throughout this section we shall be manipulating expressions pertaining to a local average of a version of the density of the SLFVFS. These manipulations are insensitive to changing that density on a set of Lebesgue measure zero and so should really be thought of as results for the measure-valued evolution itself.

Our result concerns the locally averaged process $\overline{w}_n$. However, before passage to the limit, this process is not Markov (in contrast to w itself). In order to overcome this, we follow *ETHERIDGE,*

VÉBER and YU (2019, [15]) and define

$$\phi_f(x) = \fint_{B_n(x)} f(y)\mathrm{d}y.$$

To alleviate notation, we have suppressed the dependence on n in this quantity. Notice that

$$\langle w_n, \phi_f \rangle = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} w_n(y) \frac{n^{d/3}}{V_R} f(z) \mathbf{1}_{\{|z-y| \le R_n\}}(z)\mathrm{d}y\mathrm{d}z = \langle \overline{w}_n, f \rangle.$$

Using Taylor's Theorem, we see that for $f \in C_c^2$,

$$\phi_f(x) = f(x) + \frac{n^{d/3}}{V_R} \int_{B_n(x)} (y-x) \cdot \nabla f(x)\mathrm{d}y + \mathcal{O}\left(\frac{n^{d/3}}{V_R} \int_{B_n(x) \cap S_f} |x-y|^2 \mathrm{d}x \right)$$

$$= f(x) + \mathcal{O}\left(n^{-2/3}\right)\mathbf{1}_{B_n(x) \cap S_f \ne \emptyset}, \tag{8.7}$$

where S_f denotes the support of f. Here, and throughout, we have used $|\cdot|$ to denote the Euclidean norm. This yields the useful approximation

$$\langle \overline{w}_n, f \rangle = \langle w_n, \phi_f \rangle = \langle w_n, f \rangle + \mathcal{O}\left(n^{-2/3}\right). \tag{8.8}$$

It will be convenient to have some notation.

Notation 8.3. *Suppose that immediately before a reproduction event the unscaled process takes the value w, then immediately after the event, its state will be $\theta^+_{x,R,\bar{u}}(w)$ if the parent of the event was of type a, and $\theta^-_{x,R,\bar{u}}(w)$ if the parent was of type A, where*

$$\theta^+_{x,R,\bar{u}}(w) = \mathbf{1}_{\mathbb{R}^d \backslash B(x,R)}w + \mathbf{1}_{B(x,R)}\left((1-\bar{u})w + \bar{u}\right),$$

$$\theta^-_{x,R,\bar{u}}(w) = \mathbf{1}_{\mathbb{R}^d \backslash B(x,R)}w + \mathbf{1}_{B(x,R)}(1-\bar{u})w. \tag{8.9}$$

Using this notation,

$$\left\langle \theta^+_{x,R_n,u_n}(w), \phi_f \right\rangle - \langle w, \phi_f \rangle = u_n \left\langle \mathbf{1}_{B_n(x)}(1-w), \phi_f \right\rangle, \tag{8.10}$$

$$\left\langle \theta^-_{x,R_n,u_n}(w), \phi_f \right\rangle - \langle w, \phi_f \rangle = -u_n \left\langle \mathbf{1}_{B_n(x)}w, \phi_f \right\rangle. \tag{8.11}$$

The first step will be to evaluate the generator on test functions of the form

$$\psi_{F,f}(w_n) = F\left(\langle \overline{w}_n, f \rangle\right) = F\left(\langle w_n, \phi_f \rangle\right), \tag{8.12}$$

where $F \in C^\infty(\mathbb{R})$, $f \in C_c^\infty(\mathbb{R}^d)$. Notice that these are independent of the environment ζ. Recall that the analogous functions were our starting point in Section 3. Since the support S_f of f is compact, it is only overlapped by events at a finite rate and so, writing $\mathcal{L}\psi_{F,f}$, which will be a function of the state $\zeta(\cdot)$ of the environment, for the generator of the unscaled process acting on

test functions of this form, we have

$$
\begin{aligned}
\mathcal{L}\psi_{F,f}(w,\zeta) =\ & (1-\mathbf{s})\int_{\mathbb{R}^d}\fint_{B(x,R)} w(y)\left[F(\langle\theta_{x,R,\bar{u}}^{+}(w),\phi_f\rangle)-F(\langle w,\phi_f\rangle)\right]\mathrm{d}y\mathrm{d}x \\
& +(1-\mathbf{s})\int_{\mathbb{R}^d}\fint_{B(x,R)}\{1-w(y)\}\left[F(\langle\theta_{x,R,\bar{u}}^{-}(w),\phi_f\rangle)-F(\langle w,\phi_f\rangle)\right]\mathrm{d}y\mathrm{d}x \\
& +\mathbf{s}\int_{\mathbb{R}^d}\fint_{B(x,R)^2}\left\{w(y_1)w(y_2)+\frac{1-\zeta(x)}{2}\Big[(w(y_1)(1-w(y_2))\right. \\
& \left.\qquad +(1-w(y_1))w(y_2)\Big]\right\}\times\left[F(\langle\theta_{x,R,\bar{u}}^{+}(w),\phi_f\rangle)-F(\langle w,\phi_f\rangle)\right]\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}x \\
& +\mathbf{s}\int_{\mathbb{R}^d}\fint_{B(x,R)^2}\left\{(1-w(y_1))(1-w(y_2))+\frac{1+\zeta(x)}{2}(w(y_1)(1-w(y_2))\right. \\
& \left.\qquad +(1-w(y_1))w(y_2))\right\}\times\left[F(\langle\theta_{x,R,\bar{u}}^{-}(w),\phi_f\rangle)-F(\langle w,\phi_f\rangle)\right]\mathrm{d}y_1\mathrm{d}y_2\mathrm{d}x. \\
=:\ & (1-\mathbf{s})\mathcal{L}^{neu}\psi_{F,f}(w)+\mathcal{L}^{fsel}\psi_{F,f}(w,\zeta)
\end{aligned}
$$

We recognise the generator of the neutral spatial Lambda-Fleming-Viot process (which does not depend on the environment) plus two terms reflecting the selection events. If we set the environment to be identically equal to 1, say, then the surviving selection term is precisely that considered in *ETHERIDGE, VÉBER* and *YU* (2019, [15]) in modelling selection in favour of type A individuals.

After scaling (and a change of variables), the generator of w_n acting on test functions of this form is given by

$$
\begin{aligned}
\mathcal{L}_n\psi_{F,f}(w_n,\zeta) =\ & n^{1+d/3}(1-\mathbf{s}_n)\left(\int_{\mathbb{R}^d}\overline{w}_n(x)\left[F(\langle\theta_{x,R_n,u_n}^{+}(w_n),\phi_f\rangle)-F(\langle w_n,\phi_f\rangle)\right]\mathrm{d}x\right. \\
& \left.\qquad +\int_{\mathbb{R}^d}\{1-\overline{w}_n(x)\}\left[F(\langle\theta_{x,R_n,u_n}^{-}(w_n),\phi_f\rangle)-F(\langle w_n,\phi_f\rangle)\right]\mathrm{d}x\right) \\
& +n^{1+d/3}\mathbf{s}_n\int_{\mathbb{R}^d}\left\{\overline{w}_n^2(x)+(1-\zeta(x))\overline{w}_n(x)(1-\overline{w}_n(x))\right\} \\
& \qquad\qquad\qquad\times\left[F(\langle\theta_{x,R_n,u_n}^{+}(w_n),\phi_f\rangle)-F(\langle w_n,\phi_f\rangle)\right]\mathrm{d}x \\
& +n^{1+d/3}\mathbf{s}_n\int_{\mathbb{R}^d}\left\{(1-\overline{w}_n(x))^2+(1+\zeta(x))\overline{w}_n(x)(1-\overline{w}_n(x))\right\} \\
& \qquad\qquad\qquad\times\left[F(\langle\theta_{x,R_n,u_n}^{-}(w_n),\phi_f\rangle)-F(\langle w_n,\phi_f\rangle)\right]\mathrm{d}x.
\end{aligned}
$$

By analogy with (3.3), we write this as

$$
\mathcal{L}_n\psi_{F,f}=(1-\mathbf{s}_n)\mathcal{L}_n^{neu}\psi_{F,f}+n^{\alpha}\mathcal{L}_n^{fsel}\psi_{F,f}, \tag{8.13}
$$

where

$$
\begin{aligned}
\mathcal{L}_n^{neu}\psi_{F,f}(w_n)=\ & n^{1+\frac{d}{3}}\left(\int_{\mathbb{R}^d}\overline{w}_n(x)\left[F(\langle\theta_{x,R_n,u_n}^{+}(w_n),\phi_f\rangle)-F(\langle w_n,\phi_f\rangle)\right]\mathrm{d}x\right. \\
& \left.\qquad +\int_{\mathbb{R}^d}\{1-\overline{w}_n(x)\}\left[F(\langle\theta_{x,R_n,u_n}^{-}(w_n),\phi_f\rangle)-F(\langle w_n,\phi_f\rangle)\right]\mathrm{d}x\right), \tag{8.14}
\end{aligned}
$$

and, writing $\mathbf{s}_n = n^\alpha \tilde{\mathbf{s}}_n = n^\alpha \mathbf{s}/n^{2/3}$,

$$
\mathcal{L}_n^{fsel} \psi_{F,f}(w_n, \zeta) = n^{1+d/3} \tilde{\mathbf{s}}_n \int_{\mathbb{R}^d} \{\overline{w}_n(x) - \zeta(x)\overline{w}_n(x)(1 - \overline{w}_n(x))\}
$$
$$
\times \left[F(\langle \theta_{x,R_n,u_n}^+(w_n), \phi_f \rangle) - F(\langle w_n, \phi_f \rangle) \right] \mathrm{d}x
$$
$$
+ n^{1+d/3} \tilde{\mathbf{s}}_n \int_{\mathbb{R}^d} \{(1 - \overline{w}_n(x)) + \zeta(x)\overline{w}_n(x)(1 - \overline{w}_n(x))\}
$$
$$
\times \left[F(\langle \theta_{x,R_n,u_n}^-(w_n), \phi_f \rangle) - F(\langle w_n, \phi_f \rangle) \right] \mathrm{d}x.
$$

Just as in the non-spatial case, in order to reduce the generator to the form (8.2), we appeal to the approach of *KURTZ* (1973, [28]). Entirely analogously to our calculations of Section 3 (see (3.4)), we evaluate the generator on test functions of the form

$$
G_{F,f}(w_n, \zeta) = \psi_{F,f}(w_n) + n^{-\alpha} \mathcal{L}_n^{fsel} \psi_{F,f}(w_n, \zeta), \tag{8.15}
$$

where $F \in C^\infty(\mathbb{R})$ and $f \in C_c^\infty(\mathbb{R}^d)$. The second term on the right hand side will form part of ϵ_n^f in (8.2). Evidently we must now account for the changing environment, but this is just the generator of a pure jump process in which a new state is sampled from a (mean zero) stationary distribution at the times of a Poisson process of rate $n^{2\alpha} \tau^{env}$.

Writing the full generator as $(1 - \mathbf{s}_n)\mathcal{L}_n^{neu} + n^\alpha \mathcal{L}_n^{fsel} + n^{2\alpha} \mathcal{L}_n^{env}$, as in (3.3), we observe that $\mathcal{L}_n^{env} \psi_{F,f} = 0$ (since $\psi_{F,f}$ is independent of ζ). In (8.23) below, we show that

$$
\mathcal{L}_n^{fsel} \psi_{F,f}(w_n, \zeta) = -\mathbf{s}\bar{u} V_R F'(\langle w_n, \phi_f \rangle) \int \zeta(x)\overline{w}_n(x)(1 - \overline{w}_n(x)) f(x)\mathrm{d}x + \mathcal{O}(n^{-1/3}),
$$

and since $\mathbb{E}_\pi[\zeta] = 0$, $\mathbb{E}_\pi[\mathcal{L}_n^{fsel} \psi_{F,f}] = \mathcal{O}(n^{-1/3})$. Calculating exactly as in the non-spatial case (see (3.5)), we find a spatial analogue of (3.6), that is

$$
\mathcal{L}_n G_{F,f}(w_n, \zeta) + \mathcal{O}\left(n^{-\alpha}\right) = (1 - \mathbf{s}_n)\mathcal{L}_n^{neu} \psi_{F,f}(w_n) + \mathcal{L}_n^{fsel}\left(\mathcal{L}_n^{fsel} \psi_{F,f}\right)(w_n, \zeta)
$$
$$
=: (1 - \mathbf{s}_n)\mathcal{L}_n^{neu} \psi_{F,f}(w_n) + \mathcal{G}_n \psi_{F,f}(w_n, \zeta), \tag{8.16}
$$

where we have used that $\alpha < 1/6 < 1/3$ to absorb $\mathbb{E}_\pi[\mathcal{L}_n^{fsel} \psi_{F,f}]$ into the error on the left hand side. In this notation, for $F \in C^\infty(\mathbb{R})$ and $f \in C_c^\infty(\mathbb{R}^d)$

$$
\psi_{F,f}(w_n(t)) - \int_0^t \left[(1 - \mathbf{s}_n)\mathcal{L}_n^{neu} \psi_{F,f}(w_n(s)) + \mathcal{G}_n \psi_{F,f}(w_n(s), \zeta(s)) \right] \mathrm{d}s + \mathcal{O}\left(n^{-\alpha}\right)
$$

is a martingale.

Neutral part

First we find an approximation to the part of the generator corresponding to neutral events. This mirrors the proof of Theorem 1.8 of *ETHERIDGE, VÉBER and YU* (2019, [15]) and so we recall the strategy, but omit many of the details.

Taking the expression (8.14) for $\mathcal{L}_n^{neu}$, we perform a Taylor expansion of F about $\langle w_n, \phi_f \rangle$ to obtain

$$
\int_0^t \mathcal{L}_n^{neu} \psi_{F,f}(w_n(s))\, \mathrm{d}s = \int_0^t [A_n(s) + B_n(s) + C_n(s)]\, \mathrm{d}s,
$$

where $A_n(s)$, $B_n(s)$, $C_n(s)$ can be expressed as

$$A_n(s) = u_n n^{1+\frac{d}{3}} F'\left(\langle \overline{w}_n(s), f \rangle\right) \int_{\mathbb{R}^d} \left[\overline{w}_n(s,x) \left\langle \mathbf{1}_{B_n(x)} (1 - w_n(s)), \phi_f \right\rangle \right.$$

$$\left. - (1 - \overline{w}_n(s,x)) \left\langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f \right\rangle \right] \mathrm{d}x$$

$$B_n(s) = u_n^2 n^{1+d/3} \frac{F''\left(\langle \overline{w}_n(s), f \rangle\right)}{2} \int_{\mathbb{R}^d} \left[\overline{w}_n(s,x) \left\langle \mathbf{1}_{B_n(x)} (1 - w_n(s)), \phi_f \right\rangle^2 \right.$$

$$\left. + (1 - \overline{w}_n(s,x)) \left\langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f \right\rangle^2 \right] \mathrm{d}x$$

$$C_n(s) \le \mathcal{O}(n^{-2d/3}),$$

where we have used (8.10) and (8.11) and C_n corresponds to the third order term in the Taylor expansion.

The key to controlling A_n and B_n is a Taylor expansion of ϕ_f. First we use Fubini's Theorem to write

$$A_n(s) = u_n n^{1+d/3} F'\left(\langle \overline{w}_n(s), f \rangle\right) \int_{\mathbb{R}^d} \left\{ \overline{w}_n(s,x) \langle \mathbf{1}_{B_n(x)}, \phi_f \rangle - \langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f \rangle \right\} \mathrm{d}x$$

$$= u_n \frac{n^{1+2d/3}}{V_R} F'\left(\langle \overline{w}_n(s), f \rangle\right) \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|<R_n\}} w_n(s,y) \mathbf{1}_{\{|x-z|<R_n\}} \phi_f(z) \mathrm{d}y \mathrm{d}z \right.$$

$$\left. - \int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|<R_n\}} w_n(s,y) \phi_f(y) \mathrm{d}y \right] \mathrm{d}x$$

$$= u_n n^{1+d/3} F'\left(\langle \overline{w}_n(s), f \rangle\right) \int_{\mathbb{R}^d} \frac{n^{d/3}}{V_R} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|<R_n\}} w_n(s,y) \mathbf{1}_{\{|x-z|<R_n\}}$$

$$(\phi_f(z) - \phi_f(y)) \, \mathrm{d}y \mathrm{d}z \mathrm{d}x.$$

Now Taylor-expand ϕ_f around y to obtain

$$A_n(s) = u_n n^{1+d/3} F'\left(\langle \overline{w}_n(s), f \rangle\right) \int_{\mathbb{R}^d} \left[\frac{n^{d/3}}{V_R} \int \int \mathbf{1}_{\{|x-y|<R_n\}} w_n(s,y) \mathbf{1}_{\{|x-z|<R_n\}} \right.$$

$$\times \left[D\phi_f(y)(z-y) + \frac{1}{2}(z-y)\mathbf{H}(y)\phi_f(z-y) + \mathcal{O}\left(|z-y|^3\right) \mathbf{1}_{y \in S_f} \right] \right] \mathrm{d}y \mathrm{d}z \mathrm{d}x, \quad (8.17)$$

where $\mathbf{H}$ denotes the Hessian. Using anti-symmetry of $z - y$, the term involving the gradient vanishes, as do the off-diagonal terms in the Hessian, so that (8.17) is equal to

$$u_n n^{1+d/3} F'\left(\langle \overline{w}_n(s), f \rangle\right) \int_{\mathbb{R}^d} \frac{n^{d/3}}{V_R} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|<R_n\}} \mathbf{1}_{\{|x-z|<R_n\}} w_n(s,y)$$

$$\times \frac{1}{2} \sum_{i=1}^d (z_i - y_i)^2 \frac{\partial^2}{\partial y_i^2} \phi_f(y) \mathrm{d}y \mathrm{d}z \mathrm{d}x,$$

plus a lower order term which is bounded uniformly in the time variable s by $\mathtt{Vol}(S_f) u_n C_f$ (where C_f is a bound on the third derivative of f and S_f is the support of f). The analogue of (8.7) with ϕ_f replaced by $\frac{\partial^2 \phi_f}{\partial y_i^2}$ is

$$\frac{\partial^2}{\partial y_i^2} \phi_f(y) = \frac{\partial^2 f}{\partial y_i^2}(y) + \mathcal{O}(n^{-2/3} \mathbf{1}_{B_n(y) \cap S_f \ne \emptyset}),$$

32

and (using (8.8) with Δf in place of f) we deduce that

$$
\begin{aligned}
A_n(s) &= u_n n^{1+\frac{d}{3}} F'(\langle \overline{w}_n(s), f\rangle) \int_{\mathbb{R}^d} \frac{n^{\frac{d}{3}}}{V_R} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|<R_n\}} \mathbf{1}_{\{|x-z|<R_n\}} w_n(s,y) \\
&\qquad \times \frac{1}{2} \sum_{i=1}^d (z_i - y_i)^2 \frac{\partial^2}{\partial y_i^2} f(y) \mathrm{d}y \mathrm{d}z \mathrm{d}x + \mathcal{O}(n^{-2/3}) \\
&= \frac{\overline{u}\Gamma_R}{2} F'(\langle \overline{w}_n(s), f\rangle) \int_{\mathbb{R}^d} w_n(s,y) \Delta f(y) \mathrm{d}y + \mathcal{O}(n^{-2/3}) \\
&= \frac{\overline{u}\Gamma_R}{2} F'(\langle \overline{w}_n(s), f\rangle) \langle \overline{w}_n(s), \Delta f\rangle + \mathcal{O}(n^{-2/3})
\end{aligned}
\tag{8.18}
$$

where

$$
\Gamma_R = \frac{1}{V_R} \int_{B(0,R)} \int_{B(x,R)} (z_1)^2 \mathrm{d}z \mathrm{d}x,
\tag{8.19}
$$

and z_1 is the first component of the vector $z \in \mathbb{R}^d$. The error term tends to zero uniformly in s and since $|\langle \overline{w}_n(s), f\rangle| \leq \|f\|_\infty \mathtt{Vol}(S_f)$, we have that $|A_n(s)|$ is uniformly bounded in s and n. The contribution coming from $B_n(s)$ can be treated similarly:

$$
\begin{aligned}
B_n(s) &= u_n^2 n^{1+d/3} \frac{F''(\langle \overline{w}_n(s), f\rangle)}{2} \\
&\quad \times \int_{\mathbb{R}^d} \overline{w}_n(s,x) \langle \mathbf{1}_{B_n(x)}(1-w_n(s)), \phi_f\rangle^2 + (1-\overline{w}_n(s,x)) \langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f\rangle^2 \mathrm{d}x \\
&= u_n^2 n^{1+d/3} \frac{F''(\langle \overline{w}_n(s), f\rangle)}{2} \int_{\mathbb{R}^d} \Big\{ \overline{w}_n(s,x) \langle \mathbf{1}_{B_n(x)}, \phi_f\rangle^2 \\
&\quad - 2\overline{w}_n(s,x) \langle \mathbf{1}_{B_n(x)}, \phi_f\rangle \langle w_n(s)\mathbf{1}_{B_n(x)}, \phi_f\rangle + \langle w_n(s)\mathbf{1}_{B_n(x)}, \phi_f\rangle^2 \Big\} \mathrm{d}x \\
&= u_n^2 n^{1+d/3} \frac{F''(\langle \overline{w}_n(s), f\rangle)}{2} \int_{\mathbb{R}^d} \Big\{ \overline{w}_n(s,x) \langle \mathbf{1}_{B_n(x)}, \phi_f\rangle (\langle \mathbf{1}_{B_n(x)}, \phi_f\rangle - \langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f\rangle) \\
&\quad - \langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f\rangle (\overline{w}_n(s,x) \langle \mathbf{1}_{B_n(x)}, \phi_f\rangle - \langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f\rangle) \Big\} \mathrm{d}x \\
&\leq u_n^2 n^{1+d/3} \frac{F''(\langle \overline{w}_n(s), f\rangle)}{2} \int_{S_f} 2\mathtt{Vol}(B_n(x))^2 \|f\|_\infty^2 \mathrm{d}x \leq C n^{(1-d)/3},
\end{aligned}
\tag{8.20}
$$

which implies that the sequence of quadratic variations is not only tight, but also, for dimensions $d \geq 2$, it tends to 0 uniformly on compact time sets. In other words, any randomness remaining in the limit in $d \geq 2$ will be due to the fluctuations in the environment.

In $d = 1$, using (8.7) and a further Taylor expansion to approximate $\phi_f(y)$ by $f(x)$ for $y \in B_n(x)$ (up to an error of order $n^{-1/3}$) and that $n^{1/3}u_n = \overline{u}$, $n^{1/3}\mathtt{Vol}(B_n) = V_R$, we have

$$
B_n(s) = \overline{u}^2 V_R^2 \frac{F''(\langle \overline{w}_n(s), f\rangle)}{2} \int_{\mathbb{R}} \overline{w}_n(s,x)(1-\overline{w}_n(s,x)) f(x)^2 \mathrm{d}x + \mathcal{O}(n^{-1/3}).
\tag{8.21}
$$

It is this term that (via an application of Lemma 8.4) leads to the final term in (4.4).

Fluctuating part

In order to tackle the part of the generator which describes the fluctuating selection, we need to understand the action of the operator $\mathcal{L}_n^{fsel}$ on $\mathcal{L}_n^{fsel}\psi_{F,f}(w_n, \zeta)$. It is here that we must diverge from *ETHERIDGE, VÉBER* and *YU* (2019, [15]).

As before, although the test function $\psi_{F,f}(w_n)$ does not depend on the environment, the result of applying $\mathcal{L}_n^{fsel}$ to it is a function that *does* depend on the environment, to which we must then apply $\mathcal{L}_n^{fsel}$. Recall that $\tilde{\mathbf{s}}_n = \mathbf{s}/n^{2/3}$. We take a Taylor expansion of F about $\langle w_n, \phi_f \rangle$ to obtain

$$
\mathcal{L}_n^{fsel}\psi_{F,f}(w_n, \zeta) + \mathcal{O}\left(n^{-\frac{1}{3}}\right)
$$
$$
= n^{1+\frac{d}{3}}\tilde{\mathbf{s}}_n u_n \left[\int_{\mathbb{R}^d} \{\overline{w}_n(x) - \zeta(x)\overline{w}_n(x)(1-\overline{w}_n(x))\} F'(\langle w_n, \phi_f \rangle)\langle \mathbf{1}_{B_n(x)}(1-w_n), \phi_f \rangle \mathrm{d}x \right.
$$
$$
\left. - \int_{\mathbb{R}^d} \{(1-\overline{w}_n(x)) + \zeta(x)\overline{w}_n(x)(1-\overline{w}_n(x))\} F'(\langle w_n, \phi_f \rangle)\langle \mathbf{1}_{B_n(x)}w_n, \phi_f \rangle \mathrm{d}x \right], \qquad (8.22)
$$

(plus lower order terms) where we have used (8.10) and (8.11).

Now using our scaling, (8.7) and (8.8), we obtain

$$
\mathcal{L}_n^{fsel}\psi_{F,f}(w_n, \zeta)
$$
$$
= n^{\frac{d}{3}}\mathbf{s}\bar{u}F'(\langle w_n, \phi_f \rangle)\left[\int_{\mathbb{R}^d} \{\overline{w}_n(x) - \zeta(x)\overline{w}_n(x)(1-\overline{w}_n(x))\} \right.
$$
$$
\times \left\langle \mathbf{1}_{B_n(x)}(1-w_n), f + \mathcal{O}(|x-y|)\,\mathbf{1}_{x \in S_f} \right\rangle \mathrm{d}x
$$
$$
\left. - \int_{\mathbb{R}^d} \{(1-\overline{w}_n(x)) + \zeta(x)\overline{w}_n(x)(1-\overline{w}_n(x))\} \left\langle \mathbf{1}_{B_n(x)}w_n, \left[f + \mathcal{O}(|x-y|)\,\mathbf{1}_{x \in S_f}\right] \right\rangle \mathrm{d}x \right]
$$
$$
+ \mathcal{O}(n^{-1/3})
$$
$$
= -\mathbf{s}\bar{u}V_R F'(\langle w_n, \phi_f \rangle) \int_{\mathbb{R}^d} \zeta(x)\overline{w}_n(x)(1-\overline{w}_n(x))f(x)\mathrm{d}x + \mathcal{O}(n^{-1/3})
$$
$$
= -\mathbf{s}\bar{u}V_R F'(\langle w_n, \phi_f \rangle)H_f(\overline{w}_n, \zeta) + \mathcal{O}(n^{-1/3}), \qquad (8.23)
$$

where
$$
H_f(\overline{w}_n, \zeta) = \int \zeta(x)\overline{w}_n(x)\left(1-\overline{w}_n(x)\right)f(x)\mathrm{d}x.
$$

We now need to evaluate $\mathcal{L}_n^{fsel}$ on the product in (8.23).

$$
\mathcal{L}_n^{fsel}\left(\mathcal{L}_n^{fsel}\psi_{F,f}\right)(w_n, \zeta) + \mathcal{O}(n^{-1/3})
$$
$$
= -\mathbf{s}\bar{u}V_R\mathcal{L}_n^{fsel}\left(\psi_{F',f}\right)(w_n, \zeta)H_f(\overline{w}_n, \zeta) - \mathbf{s}\bar{u}V_R F'(\langle w_n, \phi_f \rangle)\left(\mathcal{L}_n^{fsel}H_f\right)(\overline{w}_n, \zeta). \quad (8.24)
$$

As in the non-spatial case, these two terms will correspond to the diffusion and drift terms respectively in the stochastic p.d.e..

The difficulty that we now face is that H depends on $\langle \overline{w}^2, f \rangle$ as well as $\langle \overline{w}, f \rangle$ and it will no longer suffice to use the test function ϕ_f and integrate by parts. Let us consider the effect on $\overline{w}_n(x)^2$ of an event in the ball $B_n(y)$ in which the parent is of type a. After the event, $\overline{w}_n(x)^2$ is replaced by

$$
\left(\overline{w}_n(x) + \frac{n^{d/3}}{V_R} \int_{B_n(y)} \mathbf{1}_{\{|z-x|<R_n\}}u_n(1-w_n(z))\mathrm{d}z \right)^2
$$
$$
= \overline{w}_n(x)^2 + 2u_n\overline{w}_n(x)\frac{n^{d/3}}{V_R}\int_{B_n(y)} \mathbf{1}_{\{|z-x|<R_n\}}(1-w_n(z))\mathrm{d}z
$$
$$
+ \mathcal{O}(u_n^2\mathbf{1}_{\{|x-y|<2R_n\}}). \qquad (8.25)
$$

The corresponding change in $\overline{w}_n(x)(1 - \overline{w}_n(x))$ is then

$$(1 - 2\overline{w}_n(x))\frac{n^{d/3}}{V_R} \int_{B_n(y)} \mathbf{1}_{\{|z-x|<R_n\}} u_n(1 - w_n(z))\mathrm{d}z + \mathcal{O}(u_n^2 \mathbf{1}_{\{|x-y|<2R_n\}}).$$

If the parent is type A, then the change is

$$-(1 - 2\overline{w}_n(x))\frac{n^{d/3}}{V_R} \int_{B_n(y)} \mathbf{1}_{\{|z-x|<R_n\}} u_n w_n(z)\mathrm{d}z + \mathcal{O}(u_n^2 \mathbf{1}_{\{|x-y|<2R_n\}}).$$

To evaluate the second term in (8.24), we calculate the effect of an event in $B_n(y)$ and integrate with respect to y, taking into account only the first order change in $\overline{w}_n^2(x)$ as given in (8.25). We can now observe that the second term in (8.24) can be written as

$$\mathcal{L}_n^{fsel}\Big(H_f(\overline{w}_n, \zeta)\Big)$$
$$= n^{1+d/3}\tilde{\mathbf{s}}_n \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \{\overline{w}_n(y) - \zeta(y)\overline{w}_n(y)(1 - \overline{w}_n(y))\}\zeta(x)$$
$$\times \frac{n^{d/3}u_n}{V_R}\mathbf{1}_{\{|x-z|<R_n\}}\mathbf{1}_{\{|y-z|<R_n\}}(1 - w_n(z))(1 - 2\overline{w}_n(x))f(x)\mathrm{d}z\mathrm{d}y\mathrm{d}x$$
$$- n^{1+\frac{d}{3}}\tilde{\mathbf{s}}_n \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \{(1 - \overline{w}_n(y)) + \zeta(y)\overline{w}_n(y)(1 - \overline{w}_n(y))\}\zeta(x)$$
$$\times \frac{n^{d/3}u_n}{V_R}\mathbf{1}_{\{|x-z|<R_n\}}\mathbf{1}_{\{|y-z|<R_n\}}w_n(z)(1 - 2\overline{w}_n(x))f(x)\mathrm{d}z\mathrm{d}y\mathrm{d}x + \mathcal{O}(n^{-1/3}). \quad (8.26)$$

We can further simplify this expression to

$$\mathcal{L}_n^{fsel}\Big(H_f(\overline{w}_n, \zeta)\Big)$$
$$= -n^{1+d/3}\tilde{\mathbf{s}}_n \frac{n^{d/3}u_n}{V_R} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \zeta(x)\zeta(y)\mathbf{1}_{\{|x-z|<R_n\}}\mathbf{1}_{\{|y-z|<R_n\}}$$
$$\times (1 - 2\overline{w}_n(x))f(x)\overline{w}_n(y)(1 - \overline{w}_n(y))\,\mathrm{d}z\mathrm{d}y\mathrm{d}x$$
$$+ n^{1+d/3}\tilde{\mathbf{s}}_n \frac{n^{d/3}u_n}{V_R} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \Big\{\zeta(x)\mathbf{1}_{\{|x-z|<R_n\}}\mathbf{1}_{\{|y-z|<R_n\}}$$
$$\times (1 - 2\overline{w}_n(x))f(x)(\overline{w}_n(y) - w_n(z))\Big\}\mathrm{d}z\mathrm{d}y\mathrm{d}x + \mathcal{O}(n^{-1/3})$$
$$= -\bar{u}\mathbf{s}V_R \int \zeta(y)^2 \overline{w}_n(y)(1 - \overline{w}_n(y))(1 - 2\overline{w}_n(y))f(y)\mathrm{d}y$$
$$+ \bar{u}\mathbf{s}V_R \int \fint_{B_n(y)} \zeta(y)^2 2(\overline{w}_n(x) - (\overline{w}_n(y))\overline{w}_n(y)(1 - \overline{w}_n(y))f(y)\mathrm{d}y$$
$$+ \bar{u}\mathbf{s}V_R \int \fint_{B_n(y)} (\zeta(y)^2 - \zeta(x)\zeta(y))\overline{w}_n(y)(1 - \overline{w}_n(y))(1 - 2\overline{w}_n(x))f(y)\mathrm{d}x\mathrm{d}y \quad (8.27)$$
$$+ \mathcal{O}(n^{-1/3}),$$

where we have written

$$(1 - 2\overline{w}_n(x)) = (1 - 2\overline{w}_n(y)) + 2(\overline{w}_n(y) - \overline{w}_n(x)),$$

performed a Taylor expansion of f around y for $|y - x| < R_n$, and used Fubini's Theorem.

Since $F'(\langle w_n, \phi_f \rangle)$ is independent of ζ, the first term in (8.24) can be read off immediately from our previous calculations, and combining all the above, we see that we should like to take A in Theorem 8.1 to be

$$AF(\langle \cdot, f \rangle)(\overline{w}, \zeta) = F'(\langle \overline{w}, f \rangle)\left\{\frac{\overline{u}\Gamma_R}{2}\langle \overline{w}, \Delta f \rangle + \langle \overline{u}^2 \mathbf{s}^2 V_R^2 \overline{w}(1 - \overline{w})(1 - 2\overline{w}), f \rangle\right\}$$
$$+ F''(\langle \overline{w}, f \rangle)\left\{\overline{u}^2 \mathbf{s}^2 V_R^2 \langle \zeta \overline{w}(1 - \overline{w}), f \rangle^2 + \mathbf{1}_{d=1}\frac{\overline{u}^2 V_R^2}{2}\langle \overline{w}(1 - \overline{w}), f^2 \rangle\right\}, \quad (8.28)$$

and

$$\epsilon_n^f(t) = \overline{u}^2 \mathbf{s}^2 V_R^2 \int_0^t \left\{F'(\langle \overline{w}_n(s), f \rangle) \int_{\mathbb{R}^d} \fint_{B_n(y)} \left(\zeta(s, y)^2 - \zeta(s, x)\zeta(s, y)\right)\right.$$
$$\times \overline{w}_n(s, y)\left(1 - \overline{w}_n(s, y)\right)\left(1 - 2\overline{w}_n(s, x)\right) f(y)\bigg\}\mathrm{d}x\mathrm{d}y\mathrm{d}s$$
$$+ 2\overline{u}^2 \mathbf{s}^2 V_R^2 \int_0^t \left\{F'(\langle \overline{w}_n(s), f \rangle) \int_{\mathbb{R}^d} \fint_{B_n(y)} \left(\overline{w}_n(s, x) - \overline{w}_n(s, y)\right)\right.$$
$$\times \overline{w}_n(s, y)\left(1 - \overline{w}_n(s, y)\right) f(y)\bigg\}\mathrm{d}x\mathrm{d}y\mathrm{d}s + Ct\mathcal{O}(n^{-\alpha}), \quad (8.29)$$

where the term of order $n^{-\alpha}$ comes from the second term on the right of (8.15) and we have used that $\alpha < 1/6$ to absorb a term of order $n^{-1/3}$. The additional term in $AF(\langle \cdot, f \rangle)$ in $d = 1$ stems from the term $B_n(s)$ in (8.21). Since $AF(\langle \cdot, f \rangle)$ is uniformly bounded, condition (8.3) is obviously satisfied. To check that $\mathbb{E}[\sup_{t \leq T} |\epsilon_n^f(t)|] \to 0$ as $n \to \infty$, we must control the integral terms in (8.29). We shall require the following lemma.

Lemma 8.4. *Let $|y - z| < R_n$. Then the following inequality holds:*

$$\mathbb{E}[\sup_{s \leq T} |\overline{w}_n(s, y) - \overline{w}_n(s, z)|] \leq C_{\lambda, T}\left(\left(|y - z|^{1/2(d+1)} + n^{-1/6}\right)e^{\lambda|y|}\right) \quad (8.30)$$

The proof is given in Appendix A.

The expectation of the supremum of the second integral in (8.29) is immediately controlled by (8.30).

To control the first integral in (8.29), first observe that for $|x - y| < R_n$, by (4.3),

$$\mathbb{E}\left[|\zeta(y)^2 - \zeta(x)\zeta(y)|\right] = |g(y, y) - g(x, y)| \leq CR_n,$$

so that writing

$$I(s) = \int \fint_{B_n(y)} \left(\zeta(s, y)^2 - \zeta(s, x)\zeta(s, y)\right)\overline{w}_n(s, y)\left(1 - \overline{w}_n(s, y)\right)\left(1 - 2\overline{w}_n(s, x)\right) f(y)\mathrm{d}x\mathrm{d}y,$$

we have

$$\mathbb{E}_\pi\left[|I(s)|\right] \leq Cn^{-1/3}.$$

Let us write $\tau_i, i = 1, \ldots, N_T$ for the times at which the environment is resampled in the time interval $[0, T]$, and setting $\tau_0 = 0$ and $\tau_{N_T+1} = T$, write

$$S_i = \sup\{I(s) : s \in [\tau_{i-1}, \tau_i)\}, \quad i = 1, \ldots, N_T + 1.$$

By Markov's inequality, $\mathbb{P}[S_i > 1/n^\delta] \leq C n^{-1/3+\delta}$ and

$$\mathbb{P}\left[\sup_{1 \leq i \leq N_T+1} S_i > \frac{1}{n^\delta}\right] \leq \mathbb{E}\left[N_T + 1\right] \mathbb{P}\left[S_1 > \frac{1}{n^\delta}\right] \leq C n^{2\alpha-1/3+\delta} T.$$

Thus for $\delta > 0$,

$$\mathbb{E}\left[\sup_{1 \leq i \leq N_T+1} S_i\right] \leq C\left(\frac{1}{n^\delta} + \mathbb{P}\left[\sup_{1 \leq i \leq N_T+1} S_i > \frac{1}{n^\delta}\right]\right) \leq C\left(\frac{1}{n^\delta} + n^{2\alpha+\delta-1/3}\right).$$

Since $\alpha < 1/6$, we may choose $\delta > 0$ in such a way that the right hand side tends to zero, and so we conclude that for fixed T,

$$\mathbb{E}\left[\sup_{t \leq T} |\epsilon_n^f(t)|\right] \to 0$$

as $n \to \infty$, as required.

Remark 8.5. *This is the only part of the argument where the condition $\alpha < 1/6$ is required. For all other arguments $\alpha < 1/3$ would suffice.*

Remark 8.6. *The estimate on $\mathbb{E}\left[\sup_{t \leq T} |\epsilon_n^f(t)|\right]$ depends on the Lipschitz bound on the correlation kernel $g(x, y)$. This is precisely the place where our argument breaks down if we allow the environmental noise to be 'white'.*

Remark 8.7. *In our arguments we have implicitly used the fact that there exists a continuous (in time) modification of the solutions to the limiting martingale problem. Let us briefly discuss why this is the case for $d \geq 2$, the other case being completely analogous. We notice that by (4.5) we may write*

$$\langle w(t), f \rangle = M_f(t) + \langle w(0), f \rangle$$
$$+ \int_0^t \left\{\frac{\bar{u}\Gamma_R}{2}\langle w(\tau), \Delta f \rangle + \langle \bar{u}^2 \mathbf{s}^2 V_R^2 w(\tau)(1 - w(\tau))(1 - 2w(\tau)), f \rangle\right\} d\tau, \quad (8.31)$$

where $M_f(t)$ is a martingale with bracket process

$$\left[M_f\right]_t = V_R^2 \bar{u}^2 \mathbf{s}^2 \int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} g(x, y) w(\tau, x) w(\tau, y) f(x) f(y) dx dy d\tau.$$

Let $0 < s < t$ and fix $p > 2$. We may use the above representation to write

$$\mathbb{E}\left[|\langle w(t), f \rangle - \langle w(s), f \rangle|^p\right]$$
$$\leq \mathbb{E}\left[\left(\int_s^t \left\{\frac{\bar{u}\Gamma_R}{2}\langle w(\tau), \Delta f \rangle + \langle \bar{u}^2 \mathbf{s}^2 V_R^2 w(\tau)(1 - w(\tau))(1 - 2w(\tau)), f \rangle\right\} d\tau\right.\right.$$
$$\left.\left. + |M_f(t) - M_f(s)|\right)^p\right] \quad (8.32)$$

Recall that if N is a local martingale then for any $1 < p < \infty$ the Burkholder-Davis-Gundy martingale moment inequality holds:

$$\mathbb{E}\left[\sup_{0 \leq s \leq t} |N(s)|^p\right] \leq C_p \mathbb{E}\left[\left[N(t)\right]_t^{p/2}\right]. \quad (8.33)$$

Plugging (8.33) *into* (8.32) *and observing that all quantities in the integral term are uniformly bounded leads to*

$$\mathbb{E}\left[|\langle w(t), f\rangle - \langle w(s), f\rangle|^p\right] \leq C|t - s|^{p/2}$$

Since $p > 2$, the continuity (in time) follows from the Kolmogorov Continuity Theorem.

8.2 Uniqueness of solutions for the limiting equation

In order to establish uniqueness of the limit points in $d \geq 2$, and hence convergence of our rescaled SLFVFS, we consider the corresponding stochastic partial differential equation. The first task is to show that any solution to the martingale problem (4.5) is a weak solution to the stochastic p.d.e. (4.6). We establish this in Appendix B, using a technique of *KURTZ* (2010, [31]). This then reduces the question to that of the weak uniqueness of the solution to the corresponding stochastic p.d.e., which we deduce from a result of *HU, NUALART and SON* (2013, [24]).

In one dimension, although any solution to the martingale problem does indeed give a solution to the stochastic p.d.e., we do not have an analogue of the result of *HU, NUALART and SON* (2013, [24]). The only case in which we have a proof of uniqueness of the equation is when W is also space-time white noise, in which case we can use the duality of Section 5. We note, however, that the proof of our main result would need to be modified to capture this form of environmental noise, see Remark 8.6. We do not currently know how to do this.

Theorem 8.8 ([24], Theorem 4.1). *Consider the nonlinear stochastic heat equation*

$$\mathrm{d}X_t = \frac{1}{2}\Delta X_t + b(X_t)\mathrm{d}t + \sigma(X_t)W(\mathrm{d}t, \mathrm{d}x), \tag{8.34}$$

with initial data X_0, where b, σ are real-valued, Lipschitz functions and the noise W is white in time and coloured in space with quadratic variation given by

$$\langle W(\phi)\rangle_t = t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} g(x, y)\phi(x)\phi(y)\mathrm{d}x\mathrm{d}y.$$

Suppose that the correlation kernel $g(x, y)$ satisfies

$$\sup_{x\in\mathbb{R}^d} \int_0^t \int_{\mathbb{R}^{2d}} p_{t-s}(x - z_1)p_{t-s}(x - z_2)|g(z_1, z_2)|\mathrm{d}z_1\mathrm{d}z_2\mathrm{d}s < \infty, \tag{8.35}$$

where p_t is defined as

$$p_t(x) = (2\pi t)^{\frac{d}{2}} e^{-\frac{|x|^2}{2t}}.$$

Then, for any bounded initial data X_0, there exists a unique adapted process X satisfying (8.34).

Corollary 8.9. *Solutions to the stochastic p.d.e. (4.6) are unique.*

Proof. It is easy to see that all the conditions of Theorem 8.8 are satisfied. The functions b, σ, are polynomials in w^∞ and, since the solution itself take values in $[0, 1]$, they are both Lipschitz. The correlation kernel belongs to $C_b(\mathbb{R}^d \times \mathbb{R}^d)$ and so condition (8.35) is satisfied. $\qquad\square$

Remark 8.10. *Theorem 8.8 guarantees uniqueness of the mild solution of (8.34). However, under the assumptions of this Theorem, the mild and analytically weak formulations are actually equivalent; see the discussion on p. 1917 of MYTNIK, PERKINS and STURM (2006, [38]).*

Remark 8.11. *An alternative proof strategy would involve using a pathwise uniqueness result of RIPPL and STURM (2013, [44]) for a heat equation driven by coloured noise. Uniqueness of the weak solution to (4.6) would be a consequence of a Yamada-Watanabe Theorem. A suitable version of the latter is provided by KURTZ (2007, [30]) Theorem 3.14 and Lemma 2.4 (see also RIPPL and STURM (2013, [44]), Theorem 1.3, and KURTZ (2014, [32]) Example 3.9.).*

A Proof of Lemma 8.4

In this section we shall sketch the proof of Lemma 8.4. The main idea of the proof is to use the time-dependent formulation of the martingale problem to transfer the problem of estimating increments of w_n into continuity estimates of the transition densities of a certain continuous time and space random walk. We begin with some calculations which reveal the appropriate choice of test functions. First observe that setting $F(x) \equiv x$ in (8.14),

$$
\mathcal{L}_n^{neu} \psi_{\mathrm{Id},f}(w_n) = u_n n^{1+d/3} \int_{\mathbb{R}^d} \overline{w}_n(y) \int_{B_n(y)} \phi_f(x)(1 - w_n(x)) \mathrm{d}x \mathrm{d}y
$$
$$
- u_n n^{1+d/3} \int_{\mathbb{R}^d} (1 - \overline{w}_n(y)) \int_{B_n(y)} \phi_f(x) w_n(x) \mathrm{d}x \mathrm{d}y
$$
$$
= n^{(2+d)/3} \overline{u} \int \int \phi_f(x) \mathbf{1}_{\{|x-y|<R_n\}} \Big(\overline{w}_n(y) - w_n(x) \Big) \mathrm{d}x \mathrm{d}y. \quad \text{(A.1)}
$$

Let f^{*k} denote the k-fold convolution of f. Writing $\overline{w}_n$ as an integral and using the notation

$$
\sqcap_R(x) = \frac{1}{V_R} \mathbf{1}_{|x|<R},
$$

we obtain

$$
\mathcal{L}_n^{neu} \psi_{\mathrm{Id},f}(w_n) = V_R n^{2/3} \overline{u} \int_{\mathbb{R}} \phi_f(x) \left(\int \sqcap_{R_n}^{*2}(x - y) \big(w_n(y) - w_n(x) \big) \mathrm{d}y \right) \mathrm{d}x.
$$

Analogously, by setting $F(x) = x$ in (8.24) and using (8.26), we obtain

$$
\mathcal{L}_n^{fsel} \Big(\mathcal{L}_n^{fsel} \psi_{\mathrm{Id},f}(w_n) \Big) = V_R^2 \mathbf{s}^2 \overline{u}^2 \int_{\mathbb{R}^d} v_n(\zeta, w_n) \phi_f(x) \mathrm{d}x + \mathcal{O}(n^{-1/3}),
$$

where the function v_n in the integrand is a uniformly bounded function dependent on ζ and w_n (in a neighbourhood of x) whose exact value will not be important.

The main idea is to find a set of time-dependent test functions ϕ_t^n such that

$$
\mathcal{L}_n^{neu}(\langle w_n(s), \phi_n(T - s) \rangle) = 0. \quad \text{(A.2)}
$$

For this purpose, we define a family of functions $\phi_n^y(0, x) : \mathbb{R}^d \to \mathbb{R}$ by

$$
\phi_n^y(0, x) = \sqcap_{R_n}(x - y).
$$

A set of test functions which satisfy (A.2) is given by solutions to

$$
\frac{\partial}{\partial t} \phi_n^y(t, x) = (\sqcap_{R_n}^{*2} * \phi_n^y)(t, x) - \phi_n^y(t, x), \quad \text{(A.3)}
$$

with initial condition $\phi_n^y(0, x)$. Notice that the initial condition is the density of a uniform distribution over the ball of radius R_n centred at y.

With this family of test functions we consider a time-dependent martingale problem of the form

$$\langle w_n(T, \cdot), \phi_n^y(0, \cdot) \rangle$$

$$= \langle w_n(0, \cdot), \phi_n^y(T, \cdot) \rangle + M_T^{n, \phi^y} + V_R^2 \bar{u}^2 \mathbf{s}^2 \int_0^T v_n(s, x) \phi_n^y(T - s, x) \mathrm{d}x \mathrm{d}s + \mathcal{O}(n^{-\alpha})$$

where M_T^{n, ϕ_n^y} is a mean-zero martingale and we have suppressed the dependence of v_n on ζ and w_n in our notation. The question of regularity of w can therefore be translated into a question of regularity of the test function. To be a little more precise, notice that

$$\left(\overline{w}_n(T, y) - \overline{w}_n(T, z) \right) = \frac{1}{V_{R_n}} \langle w_n(T, \cdot), \phi_n^y(0, \cdot) \rangle - \langle w_n(T, \cdot), \phi_n^z(0, \cdot) \rangle$$

$$= \langle w_n(T, \cdot), \phi_n^y(0, \cdot) - \phi_n^z(0, \cdot) \rangle$$

and therefore, using the time dependent martingale problem,

$$\left(\overline{w}_n(T, y) - \overline{w}_n(T, z) \right)$$

$$= \langle w_n(0, \cdot), \phi_n^y(T, \cdot) - \phi_n^z(T, \cdot) \rangle + M_T^{n, \phi^y} - M_T^{n, \phi^z}$$

$$+ \int_0^T \int_{\mathbb{R}^d} v_n(s, x) \left(\phi_n^y(T - s, x) - \phi_n^z(T - s, x) \right) \mathrm{d}x \mathrm{d}s + \mathcal{O}(n^{-\alpha}). \quad \text{(A.4)}$$

Remark A.1. *We have defined the generator of the process for functions which are twice differentiable in the spatial coordinates. The characteristic function of a ball clearly does not satisfy this assumption. However, we may approximate such a function with a smooth function, making sure that the approximation does not lead to an error of order greater than $n^{-1/3}$. We shall ignore this error in what follows.*

Observe that ϕ_n^y has a probabilistic interpretation. It is the density of the continuous time random walk with jump rate $n^{2/3}$ for which the distribution of an increment is given by that of the sum of two independent uniform random variables on $B_{R_n}(0)$, and for which the initial position is a point picked uniformly from the ball of radius R_n about y.

It will be convenient to write $\Lambda_{R_n} = \sqcap_{R_n}^{*2}$. The density of the random walk above, when started from the point y, which we shall denote by ψ_n^y, can be written as

$$\psi_n^y(t, x) = \sum_{k=0}^{\infty} e^{-n^{2/3}t} \frac{(n^{2/3}t)^k}{k!} \sqcap_{R_n}^{*2k}(x - y) = \sum_{k=0}^{\infty} e^{-n^{2/3}t} \frac{(n^{2/3}t)^k}{k!} \Lambda_{R_n}^{*k}(x - y). \quad \text{(A.5)}$$

As n tends to infinity, ψ_n tends to the density of a Brownian motion.

We shall require the following Lemma, which is an analogue of Lemma 3 of *MUELLER and TRIBE* (1995, [36]).

Lemma A.2. *Let $\|f\|_\lambda = \sup_x |f(x)| e^{\lambda|x|}$. Then, for any $0 < t < T$,*

i) For any $y \in \mathbb{R}^d$,

$$\left| \psi_n^y(t, x) - e^{-n^{2/3}t} \delta_y(x) - \sum_{k=1}^{\infty} e^{-n^{2/3}t} \frac{(n^{2/3}t)^k}{k!} p\left(\frac{k}{n^{2/3}} \Gamma, x - y \right) \right| \leq C_T n^{-1/3} t^{-1/2},$$

where

$$p(t,z) := \frac{1}{(2\pi t)^{d/2}} \exp\left(-|z|^2/2t\right),$$

and the constant Γ was defined in (8.19);

ii) For any $y, z \in \mathbb{R}^d$,

$$|\psi_n^y(t,x) - \psi_n^z(t,x) - e^{-n^{2/3}t}\delta_y(x) + e^{-n^{2/3}t}\delta_z(x)|$$
$$\leq C_T\left(t^{-(d+1)/2}|y-z| + n^{-1/3}t^{-1/2}\right);$$

iii) For $|x - y| > 1$ and $\lambda > 0$,

$$\psi_n^y(t,x) < C_{\lambda,T} e^{-\lambda|x-y|};$$

iv) For $|y - z| < 1$ and $\lambda > 0$,

$$\|\phi_n^y(t,x) - \phi_n^z(t,x) - e^{-n^{2/3}t}\sqcap_{R_n}(x-y) + e^{-n^{2/3}t}\sqcap_{R_n}(x-z)\|_\lambda$$
$$\leq C_{\lambda,d,T} e^{\lambda|y|}\left(t^{-1/2}|y-z|^{1/(d+1)} + n^{-1/3}t^{-1/2}\right).$$

We defer a sketch of the proof of this Lemma until the end of this Appendix. Armed with Lemma A.2, we may now sketch the proof of (8.30). Since all parts of the estimates have a similar flavour, we focus on the integral term of (A.4). As above, we shall write $v_n(s,x)$ for a uniformly bounded function, whose exact value, which depends on ζ and $w_n(s,\cdot)$ in a neighbourhood of x, will not be important. With the intention of using the bounds in Lemma A.2 we write

$$\left|\int_0^T \int_{\mathbb{R}^d}\left[\phi_n^y(T-s,x) - \phi_n^z(T-s,x)\right]v_n(s,x)\mathrm{d}x\mathrm{d}s\right| \leq I_1 + I_2 + I_3$$

where

$$I_1 = \left|\int_0^{T-\tau}\int_{\mathbb{R}^d}\left[\phi_n^y(T-s,x) - \phi_n^z(T-s,x)\right]v_n(s,x)\mathrm{d}x\mathrm{d}s\right.$$
$$\left. - \int_0^{T-\tau}\int_{\mathbb{R}^d} e^{-n^{2/3}(t-s)}v_n(s,x)\left[\phi_n^y(0,x) - \phi_n^z(0,x)\right]\mathrm{d}x\mathrm{d}s\right|,$$

$$I_2 = \left|\int_{T-\tau}^T\int_{\mathbb{R}^d}\left[\phi_n^y(T-s,x) - \phi_n^z(T-s,x)\right]v_n(s,x)\mathrm{d}x\mathrm{d}s\right.$$
$$\left. - \int_{T-\tau}^T\int_{\mathbb{R}^d} e^{-n^{2/3}(T-s)}v_n(s,x)\left[\phi_n^y(0,x) - \phi_n^z(0,x)\right]\mathrm{d}x\mathrm{d}s\right|,$$

$$I_3 = \left|\int_0^T\int_{\mathbb{R}^d} e^{-n^{2/3}(T-s)}v_n(s,x)\left[\phi_n^y(0,x) - \phi_n^z(0,x)\right]\mathrm{d}x\mathrm{d}s\right|.$$

We estimate these three terms separately. First observe that

$$I_1 = \left| \int_0^{T-\tau} \int_{\mathbb{R}^d} \left[\phi_n^y(T-s,x) - \phi_n^z(T-s,x) \right. \right.$$
$$\left. \left. + e^{-n^{2/3}(T-s)} \big(\phi_n^y(0,x) - \phi_n^z(0,x) \big) \right] v_n(s,x) \mathrm{d}x \mathrm{d}s \right|$$
$$\leq \sup_{s \in [0,T-\tau]} \| \phi_n^y(T-s,x) - \phi_n^z(T-s,x) + e^{-n^{2/3}(T-s)} \big(\phi_n^y(0,x) - \phi_n^z(0,x) \big) \|_\lambda$$
$$\times \int_0^{T-\tau} \int_{\mathbb{R}^d} e^{-\lambda|x|} |v_n(s,x)| \mathrm{d}x \mathrm{d}s$$
$$\leq C_{\lambda,T} \left(\tau^{-1/2} |y-z|^{1/(d+1)} + n^{-1/3} \tau^{-1/2} \right) e^{\lambda|y|},$$

where we have used Part iv) of Lemma A.2. Analogously

$$I_2 = \left| \int_{T-\tau}^T \int_{\mathbb{R}^d} \left[\phi_n^y(T-s,x) - \phi_n^z(T-s,x) \right. \right.$$
$$\left. \left. + e^{-n^{2/3}(T-s)} \big(\phi_n^y(0,x) - \phi_n^z(0,x) \big) \right] v_n(s,x) \mathrm{d}x \mathrm{d}s \right|$$
$$\leq C_\lambda \int_{T-\tau}^T \| \phi_n^y(T-s,x) - \phi_n^z(T-s,x) + e^{-n^{2/3}(T-s)} \big(\phi_n^y(0,x) - \phi_n^z(0,x) \big) \|_\lambda \mathrm{d}s$$
$$\leq C_{\lambda,t} \tau^{1/2} e^{\lambda|y|},$$

where we have again used Part iv) of Lemma A.2. Finally, I_3 can be bounded by

$$I_3 \leq C \int_0^T e^{-n^{2/3}s} \mathrm{d}s \leq C n^{-2/3}.$$

Combining the estimates above, and observing that the term which involves $w_n(0,\cdot)$ can be tackled analogously, we conclude that for $|y-z| < 1$,

$$|\overline{w}_n(t,y) - \overline{w}_n(t,z)|$$
$$\leq \left| M_t^{n,\phi_n^y} - M_t^{n,\phi_n^z} \right| + |\langle w_n(0,\cdot), \phi_n^y(T,\cdot) - \phi_n(T,z,\cdot)\rangle|$$
$$+ \left| \int_0^t \int_{\mathbb{R}} \{ \phi_n^y(T-s,x) - \phi_n^z(T-s,x) \} v_n(s,x) \mathrm{d}x \mathrm{d}s \right|$$
$$\leq \left| M_t^{n,\phi^y} - M_t^{n,\phi^z} \right| + C \left(|y-z|^{1/(d+1)} \tau^{-1/2} + n^{-1/3} \tau^{-1/2} \right) e^{\lambda|y|}. \quad \text{(A.6)}$$

Choosing $\tau = \max\{|y-z|^{1/(d+1)}, n^{-1/3}\}$, the deterministic term on the right hand side is at most

$$C \left(|y-z|^{1/2(d+1)} + n^{-1/6} \right) e^{\lambda|y|}.$$

It remains to bound the martingale difference. We use the Burkholder-Davis-Gundy inequality in the form

$$\mathbb{E} \left[\sup_{s \leq t} M_s \right] \leq C \left(\mathbb{E} \langle M \rangle_t^{1/2} + \sup_{s \leq t} |M_s - M_{s_-}| \right),$$

see *HALL and HEYDE*, (1980,[22]), Theorem 2.11 for the proof in the discrete time setting, the continuous time case follows by approximation.

The first term can be controlled in the same way as above, by writing it as an integral of $(\phi_n^y - \phi_n^z)$ times a bounded function and the jumps of the martingale are of size $\mathcal{O}(n^{-1/3})$, so combining with (A.6) leads to, for any $\lambda > 0$ and $|y - z| < 1$,

$$\mathbb{E}[\sup_{s \leq T} |\overline{w}_n(s, y) - \overline{w}_n(s, z)|] \leq C_{\lambda,T} \left(|y - z|^{1/2(d+1)} + n^{-1/6} \right) e^{\lambda|y|}, \tag{A.7}$$

which concludes the proof of the Lemma. □

Proof of Lemma A.2. To prove the first part of Lemma A.2, we first observe that if $\mathcal{P}_t$ is a Poisson process with rate $n^{2/3}$, then for $a > 0$,

$$\mathbb{E}[\mathcal{P}_t^{-a} \mathbf{1}_{\mathcal{P}_t \geq 1}] \leq C_a (n^{2/3} t)^{-a}. \tag{A.8}$$

Next we use the Local Central Limit Theorem for discrete time random walk to approximate the summands in (A.5) for $k \geq 1$. To this end, let $\{X_i\}_{i \geq 1}$ be independent random variables with density Λ_{R_1}. Then by the local Central Limit Theorem, see e.g. Theorem 3 of PETROV (1964,[41]),

$$\mathbb{P}\left[\frac{1}{\sqrt{k}} \sum_{i=1}^{k} X_i = x \right] = p(\Gamma, x) + e_k(x), \tag{A.9}$$

where

$$p(t, z) := \frac{1}{(2\pi t)^{d/2}} \exp\left(-|z|^2/2t \right),$$

and the error e_k is bounded by $C/\sqrt{k}$ with C independent of k and x.

Now using that the $\Lambda_{R_n}^{*k}$ in the kth summand in (A.5) is the density function of $\sqrt{k}/n^{1/3}$ times the random variable in (A.9), a change of variable allows us to express it as $p(\Gamma k/n^{2/3}, x - y)$ plus an error of order $(k/n^{2/3})^{d/2} 1/\sqrt{k}$.

To estimate the error term in i), we split the sum in (A.5) into two parts, $\{k \leq 2eTn^{2/3}\}$ and $\{k > 2eTn^{2/3}\}$. For the first, $(k/n^{2/3})^{d/2}$ is uniformly bounded and exploiting (A.8) with $a = 1/2$ yields a term of the form of the error in i).

To control the remaining term when $k > eTn^{2/3}$, first observe that by a Chernoff bound for the Poisson distribution,

$$\mathbb{P}[\mathcal{P}_t > 2eTn^{2/3}] \leq \left(\frac{t}{2T} \right)^{2eTn^{2/3}} e^{-n^{2/3}t}.$$

Now following the proof of Lemma B.5 of *ETHERIDGE, VÉBER and YU* (2019, [15]), first observe that if Z_R is a uniform random variable on $B(0, R) \subset \mathbb{R}^d$, then

$$\tilde{\rho}_R(\theta) := \mathbb{E}[e^{i\theta . Z_R}] = \frac{2^{d/2} \Gamma(d/2 + 1)}{|R\theta|^{d/2}} J_{d/2}(|R\theta|),$$

where $J_{d/2}$ is the Bessel function of the first kind of order $d/2$. Moreover, it is shown there that for $|\theta| > n^{1/3}$ there is $a \in (0, 1)$ with $\tilde{\rho}_R(\theta)^2 \leq a$, from which we see that if $|\theta| > n^{1/3}$, $\tilde{\rho}_{R/n^{1/3}}(\theta)^2 \leq a$. Indeed, since $|J_\nu(z)| \leq C_\nu z^{-1/2}$ for $\nu > 0$, we find that for $|\xi| > n^{1/3}$, $\tilde{\rho}(R_n \xi)^2 \leq a \wedge C_d |\xi|^{-(d+1)}$.

For $|\xi| \leq n^{1/3}$, we use the bound $|\tilde{\rho}_{R_n}(\xi)| \leq 1$.

Inverting the Fourier transform, from these bounds we find

$$\sum_{k=2eTn^{2/3}}^{\infty} e^{-n^{2/3}t} \frac{(n^{2/3}t)^k}{k!} \Lambda_{R_n}^{*k}(x-y) \leq \left(\frac{t}{2T}\right)^{2eTn^{2/3}} e^{-n^{2/3}t} C\left(n^{d/3} + a^{2eTn^{2/3}-1}\right). \quad (A.10)$$

Evidently this can also be absorbed into the error term. An identical argument controls the part of the sum over normal densities corresponding to $k > 2eTn^{2/3}$, and the proof of i) is complete.

For ii), we use i) to approximate both terms by sums of normal densities plus an error of order $n^{-1/3}t^{-1/2}$. Now observe that

$$|p(t,y) - p(t,z)| = t^{-d/2}|p(1, t^{-1/2}y) - p(1, t^{-1/2}z)|$$
$$\leq Ct^{-d/2}|t^{-1/2}x - t^{-1/2}y| \leq Ct^{-(d+1)/2}|y-z|,$$

and sum using the estimate (A.8) to conclude this part of the proof.

For $iii)$, let S_k be a random variable with density $\Lambda_{R_n}^{*k}$. Writing $x = (x_1, x_2, \ldots, x_d)$, we observe that since S_k is symmetric, it suffices to prove the result with $x = (x_1, 0, \ldots, 0)$ and $x_1 > 1$. Since the density is unimodal and we are assuming that $|x| > 1$, $\mathbb{P}[S_k = (x_1, 0, \ldots, 0)] \leq \mathbb{P}[S_k^{(1)} > |x| - 1]$, where we have used $S_k^{(1)}$ to denote the first component of S_k.

$$\Lambda^{*k}((x_1, 0, \ldots, 0)) \leq \mathbb{P}[S_k^{(1)} > |x| - 1] \leq e^{-\lambda(|x|-1)} \mathbb{E}\left[e^{\lambda S_k^{(1)}}\right] = \mathbb{E}\left[e^{\lambda S_1^{(1)}}\right]^k. \quad (A.11)$$

The quantity on the right is maximised in $d = 1$, when

$$\mathbb{E}[e^{\lambda S_1}] = \frac{e^{2\lambda/n^{1/3}} + e^{-2\lambda/n^{1/3}} - 2}{4\lambda^2/n^{2/3}} \leq e^{4\lambda^2/n^{2/3}}.$$

Substituting into (A.5), we find

$$\psi_n^y(t,x) \leq e^{-\lambda(|x|-1)} \exp\left(n^{2/3}t\left(e^{4\lambda^2/n^{2/3}} - 1\right)\right) \to e^{-\lambda(|x|-1)}e^{4\lambda^2} \qquad \text{as } n \to \infty,$$

which completes the proof of iii).

For $iv)$, we begin by observing that

$$|\phi_n^y(t,x) - \phi_n^z(t,x)| = \int_{\mathbb{R}^d} \left[\phi_n^0(0, x'-y) - \phi_n^0(0, x'-z)\right] \psi_n(x-x') \mathrm{d}x'$$
$$= \int_{\mathbb{R}^d} \phi_n^0(0, x') \left[\psi_n^0(x-y-x') - \psi_n^0(x-z-x')\right] \mathrm{d}x' \quad (A.12)$$

We focus on bounding the supremum of the difference of ψ. Recall that if $a, b > 0$, $\min(a,b) \leq a^{1/(d+1)}b^{d/(d+1)}$. We use $ii)$ and $iii)$ to bound

$$\sup_x (\psi_n^0(t, y-x) - \psi_n^0(t, z-x))e^{\lambda|x|}$$

$$\leq \sup_{|x-y|\leq 2} (\psi_n^0(t, y-x) - \psi_n^0(t, z-x))e^{\lambda|x|} + \sup_{|x-y|>2} (\psi_n^0(t, y-x) - \psi_n^0(t, z-x))e^{\lambda|x|}$$

$$\leq C_{\lambda,d,T}\left(t^{-(d+1)/2}|y-z| + n^{-1/3}t^{-1/2}\right)e^{\lambda|y|}$$

$$+ \sup_{|x-y|>2} \min\left(C_{\lambda,d,T}\left(t^{-(d+1)/2}|y-z| + n^{-1/3}t^{-1/2}\right)e^{\lambda|x|}, \left(e^{-2\lambda(x-y)} + e^{-2\lambda(x-z)}\right)e^{\lambda|x|}\right)$$

$$\leq C_{\lambda,d,T}\left(t^{-1/2}|y-z|^{1/(d+1)} + n^{-1/3}t^{-1/2}\right)e^{\lambda|y|}$$

The result follows immediately from (A.12). $\qquad \square$

B Equivalence of martingale problem and stochastic p.d.e. formulations

Suppose that $d \geq 2$. Write w for a continuous in time representative of the density of the $\mathcal{M}_\lambda$-valued process, which is characterised as a solution to the martingale problem (4.5). Our aim is to show that w is also a weak solution to the stochastic p.d.e. (4.6). This follows as a special case from the following result.

Theorem B.1. *Let* b, σ, g *be functions satisfying the conditions of Theorem 8.8. Suppose that for all* $F \in C^\infty(\mathbb{R})$ *and* $f \in C_c^\infty(\mathbb{R}^d)$,

$$F(\langle w_t, f \rangle) - F(\langle w_0, f \rangle) - \int_0^t F'(\langle w_s, f \rangle) \left\{ \frac{1}{2} \langle w_s, \Delta f \rangle + \langle b(s, w_s), f \rangle \right\} \mathrm{d}s \tag{B.1}$$

is a continuous martingale with quadratic variation

$$\int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} F''(\langle w_s, f \rangle) g(x, y) \sigma(s, x, w_s) \sigma(s, y, w_s) f(x) f(y) \mathrm{d}x \mathrm{d}y \mathrm{d}s.$$

Then w_t *is a (stochastically and analytically) weak solution to the stochastic p.d.e.*

$$dw_t = \left\{ \frac{1}{2} \Delta w_t + b(t, x, w_t) \right\} dt + \sigma(t, x, w_t) W(\mathrm{d}t, \mathrm{d}x), \tag{B.2}$$

where the noise W *is white in time and coloured in space with quadratic variation*

$$\langle W(\phi) \rangle_t = t \int_{\mathbb{R}^d \times \mathbb{R}^d} \phi(x) \phi(y) g(x, y) \mathrm{d}x \mathrm{d}y.$$

The converse of Theorem B.1 is a simple application of Itô's formula.

We closely follow *KURTZ* (2010, [31]), who provides a powerful proof of the corresponding result for stochastic (ordinary) differential equations, not through the classical approach of explicitly constructing the driving noise, but via the Markov Mapping Theorem. The proof of our case requires only very minor modifications of the proof from *KURTZ* (2010, [31]). This proof is very flexible and could be adapted further without any major difficulties to cover a wider class of equations, e.g. with more singular coefficients b, σ in (B.2). For convenience, we first recall the Markov Mapping Theorem in the form stated there.

Theorem B.2 (Markov Mapping Theorem, *KURTZ* (2010, [31]), Theorem 1.4). *Suppose that* E *is a complete separable metric space and that the operator* $B \subseteq \overline{C}(E) \times \overline{C}(E)$ *is separable and a pre-generator and that its domain* $\mathcal{D}(B)$ *is closed under multiplication and separates points in* E. *Let* (E_0, r_0) *be a complete, separable, metric space and* $\gamma : E \to E_0$ *be Borel measurable. Let* $\mathcal{P}(E)$ *denote the space of probability measures over* E. *Let* α *be a transition function from* E_0 *into* E *(that is, assume that* $y \in E_0 \to \alpha(y, \cdot) \in \mathcal{P}(E)$ *is Borel measurable) satisfying* $\alpha(y, \gamma^{-1}(y)) = 1$. *Define*

$$C = \left\{ \left(\int_E f(z) \alpha(\cdot, dz), \int_E Bf(z) \alpha(\cdot, dz) \right) : f \in \mathcal{D}(B) \right\}.$$

Let $\mu_0 \in \mathcal{P}(E_0)$, *and define* $\nu_0 = \int \alpha(y, \cdot) \mu_0(dy)$. *If* $\widetilde{U}$ *is a solution of the martingale problem for* (C, μ_0), *then there exists a solution* V *of the martingale problem for* (B, ν_0) *such that* $\widetilde{U}$ *has*

the same distribution on $M_{E_0}[0,\infty)$ (the space of measurable functions mapping $[0,\infty)$ to E_0, with topology given by convergence in Lebesgue measure) as $U = \gamma \circ V$ and

$$\mathbb{P}\left[V(t) \in \Gamma | \widehat{\mathcal{F}}_t^U\right] = \alpha(U(t),\Gamma), \quad \Gamma \in \mathcal{B}(E), t \in T^U, \tag{B.3}$$

where $\widehat{\mathcal{F}}_t^U$ is the completion of $\sigma\left(\int_0^r h(U(s))\mathrm{d}s : r \leq t) \vee \sigma(U_0)\right)$ (for bounded and measurable h) and T^U is the set of times for which $U(t)$ is $\widehat{\mathcal{F}}_t^U$-measurable.

The proof of Theorem B.1 is based on the introduction of an auxiliary process Z, whose generator plays the role of B in the statement of Theorem B.2. This process is introduced in such a way that we can evaluate the conditional distributions of the driving noise given the state of the process w. Those conditional distributions are given in terms of stationary processes.

First we describe the representation of the driving noise in terms of a countable collection of stationary processes. By our assumptions, the noise can be realised as an element of the Hilbert space $W^{-1-d,2}(\mathbb{R}^d)$ (see, e.g. RIPPL (2012, [43])), that is on the dual of the Sobolev space $W^{1+d,2}(\mathbb{R}^d)$. Therefore there exists an orthonormal basis $\{\phi_i\}_{i\geq 1}$ such that the noise W is completely characterised by a countable sequence of independent, one-dimensional Brownian motions $W_t(\phi_i)$ with

$$W_t(\phi_i) = \int_0^t \int_{\mathbb{R}^d} \phi_i(x)W(ds,dx),$$

and, for any adapted process H, the integral with respect to the noise can be written in the form

$$\int_0^t \int_{\mathbb{R}^d} H(s_-,x)W(ds,dx) = \sum_{i=1}^\infty \int_0^t \langle H(s_-,\cdot),\phi_i\rangle \mathrm{d}W_s(\phi_i).$$

Define $Y = \{Y_i\}_{i\geq 1}$ by

$$Y_i(t) = \left(Y_i(0) + W_t(\phi_i) \mod 2\pi\right).$$

We can recover $W_t(\phi_i)$ (and hence W) from Y_i, by observing the increments. Indeed, setting

$$\beta_i(t) = \begin{pmatrix} \cos(Y_i(t)) + \frac{1}{2}\int_0^t \cos(Y_i(s))ds \\ \sin(Y_i(t)) + \frac{1}{2}\int_0^t \sin(Y_i(s))ds \end{pmatrix},$$

we have

$$W_t(\phi_i) = \int_0^t \left(-\sin(Y_i(s)), \cos(Y_i(s))\right) d\beta_i(s) \tag{B.4}$$

and

$$d\beta_i(t) = \begin{pmatrix} -\sin(Y_i(t)) \\ \cos(Y_i(t)) \end{pmatrix} dW_t(\phi_i).$$

Remark B.3. *Our definition ensures that $Y_i(t)$ are independent. Identifying 0 and 2π, we note that $Y = \{Y_i\}_{i\geq 1}$ is a Markov process with compact state space $[0,2\pi)^\infty$. If $Y_i(0)$ is uniformly distributed on $[0,2\pi)$ and independent of W, then Y_i is stationary.*

We define an auxiliary process Z by

$$Z = \begin{pmatrix} w \\ Y \end{pmatrix},$$

where w is a solution of the martingale problem (B.1) and $Y = \{Y_i\}_{i \geq 1}$. Let A be the generator associated with the martingale problem (B.1). Its domain is given by

$$\mathcal{D}(A) = \left\{ F(\langle \cdot, f \rangle) : f \in C_c^\infty(\mathbb{R}^d), F \in C^\infty(\mathbb{R}) \right\}.$$

Let $D_0([0, 2\pi))$ be the collection of twice differentiable functions $f : [0, 2\pi) \to [0, 2\pi)$ for which

$$f(0) = f(2\pi_-), \quad f'(0) = f'(2\pi_-), \quad f''(0) = f''(2\pi_-).$$

Let $\widehat{A}$ be the generator of the process Z. Its domain $\mathcal{D}(\widehat{A})$ is defined by

$$\mathcal{D}(\widehat{A}) = \left\{ g(w, y) := \psi(w) \prod_{i=1}^m f_i(y_i) : \psi \in \mathcal{D}(A), f_i \in D_0\big([0, 2\pi)\big), m \in \mathbb{N} \right\}.$$

An application of Itô's formula guarantees that for every $g \in \mathcal{D}(\widehat{A})$

$$\widehat{A}g(w, y) = \prod_{i=1}^m f_i(y) A\psi(w) + \psi(w) \sum_{i=1}^m \mathcal{L}_i f_i(y_i) \prod_{j=1, j \neq i}^m f_j(y_j) + \sum_{i=1}^m c_i(w) \partial_{y_i} f_i(y_i),$$

where $\mathcal{L}_i$ is the generator of Y_i and the last term (which is finite due to our construction of the noise) is determined by the Meyer process between w and Y_i. The exact value of the c_i's will not concern us. We write it in this form to stress the dependence on the derivative of f_i.

To show that solutions to the martingale problem for $\widehat{A}$ are weak solutions to (B.2), we use the following lemma.

Lemma B.4 (*KURTZ (2010, [31]), Lemma A.1*)*.* *Suppose that E is a complete, separable metric space. Let $A \subset B(E) \times B(E)$ be a generator, and let X be a càdlàg solution of the martingale problem for A. For each $f \in \mathcal{D}(A)$, define*

$$N_f(t) = f(X_t) - \int_0^t Af(X_s)\mathrm{d}s.$$

Suppose $\mathcal{D}(A)$ is an algebra and that $f \circ X$ is càdlàg for each $f \in \mathcal{D}(A)$. Let $f_0 \cdots, f_m \in \mathcal{D}(A)$ and $h_0, \cdots, h_m \in B(E)$. Then

$$N(t) = \sum_i \int_0^t h_i(X_{s_-})\mathrm{d}N_{f_i}(s)$$

is a square integrable martingale with Meyer process

$$\langle N \rangle_t = \sum_{i,j} \int_0^t h_i(X_s)h_j(X(s))$$

$$\times \left[Af_i f_j(X_s) - f_i(X_s)Af_j(X_s) - f_j(X_s)Af_i(X_s) \right]\mathrm{d}s \quad \text{(B.5)}$$

Lemma B.5. *Let $Z = (w, Y)$ be a continuous solution to the martingale problem for $\widehat{A}$. Then w is a weak solution to the stochastic p.d.e. (B.2).*

Proof. The statement follows from an application of Lemma B.4 with

$$f_0(w, \bar{y}) = F(\langle w, f \rangle), \quad f_i(w, \bar{y}) = \cos(y_i) \text{ for } i \in \{1, \dots m\},$$
$$f_i(w, \bar{y}) = \sin(y_i) \text{ for } i \in \{m+1, \dots 2m\},$$
$$h_0(w, \bar{y}) = 1, \quad h_i(w, \bar{y}) = c_i(w) \sin(y_i) \text{ for } i \in \{1, \dots m\},$$
$$h_i(w, \bar{y}) = -c_i(w) \cos(y_i) \text{ for } i \in \{m+1, \dots 2m\},$$

where we have used $\bar{y}$ to denote the vector $(y_i)_{i \geq 1}$. This choice of functions guarantees that $N(t)$ defined as in Lemma B.4 is a continuous martingale with quadratic variation equal to 0. This implies that $N(t) = 0$. Using (B.4) and rearranging, we observe that this implies that

$$F(\langle w_t, f \rangle) = F(\langle w_0, f \rangle) + \int_0^t F'(\langle w_s, f \rangle) \left\{ \frac{1}{2} \langle w_s, \Delta f \rangle + \langle b(s, w_s), f \rangle \right\} \mathrm{d}s$$
$$+ \int_0^t F''(\langle w_s, f \rangle) \langle \sigma(s, x, w_s), f \rangle W(\mathrm{d}t, \mathrm{d}x).$$

This, in turn, implies that w is a weak solution to (B.2). $\qquad\square$

Proof of Theorem B.1. We begin by observing that $D(\widehat{A})$ is closed under multiplication and separable. Separability of the operator follows from separability of its domain. The operator $\widehat{A}$ is a pre-generator and separates points. We proceed to construct the transition function α. Let $\eta_y \in \mathcal{P}([0, 2\pi)^\infty)$ be the product of independent uniform distributions on $[0, 2\pi)$. We define

$$\alpha(w, \cdot) = \delta_w \times \eta_y \in \mathcal{P}\left(C_b(\mathbb{R}^d) \times [0, 2\pi)^\infty\right).$$

The function γ can be defined as a projection into the first coordinate, that is $\gamma(x, y) = x$.

We observe that
$$\alpha L_x f(x, y) = L_x \alpha f(x, y). \tag{B.6}$$

Since η_y has been chosen in a way which guarantees that it is a stationary distribution for Y, we have, by definition,
$$\alpha L_y f(x, y) = 0. \tag{B.7}$$

Since $D_0([0, 2\pi)$ is a collection of periodic functions, integration by parts shows that
$$\alpha L_{xy} f(x, y) = 0. \tag{B.8}$$

Combining (B.6), (B.7), (B.8),
$$\alpha \widehat{A} f(x, y) = A \alpha f(x, y).$$

The result follows by application of the Markov Mapping Theorem with $C = A$ and $B = \hat{A}$. $\qquad\square$

C   A brief discussion of topology

This Appendix is dedicated to description of the state space of the spatial Lambda-Fleming-Viot model and the topology with which we are working. This discussion is reminiscent of Section 2 of *VÉBER and WAKOLBINGER*. We begin by recalling the definition of the space. Let $\mathcal{K} = \{a, A\}$ and let $C_c(\mathbb{R}^d)$ denote the space of all compactly supported continuous functions on $\mathbb{R}^d$. The space

$\mathcal{M}_\lambda(\mathbb{R}^d \times \mathcal{K})$ is the space of all measures on $\mathbb{R}^d \times \mathcal{K}$ with the first component being equal to Lebesgue measure. More formally,

$$\mathcal{M}_\lambda := \left\{ M \in \mathcal{M}(\mathbb{R}^d \times \mathcal{K}) : \int_{\mathbb{R}^d \times \mathcal{K}} f(x) M(\mathrm{d}x \mathrm{d}\kappa) = \int_{\mathbb{R}^d} f(x)\mathrm{d}x \quad \forall f \in C_c(R^d) \right\}.$$

An application of the disintegration theorem leads to existence of a density $w : \mathbb{R}^d \to [0,1]$. The density can be recovered from the formula

$$M_t(\mathrm{d}x, \mathrm{d}\kappa) = w(t,x)\delta_a(\mathrm{d}\kappa) + (1 - w(t,x))\delta_A(\mathrm{d}\kappa)$$

and is unique up to sets of Lebesgue measure 0. The usual choice of topology on $\mathcal{M}_\lambda$ is the topology of vague convergence. To check that the sequence of measures $M_n \in \mathcal{M}_\lambda$ with densities w_n converges vaguely to M with density w, it is sufficient to show that for every $f \in C_c(\mathbb{R}^d)$

$$\int w_n(x) f(x)\mathrm{d}x \to \int w(x) f(x)\mathrm{d}x.$$

As Theorem 8.1 requires the continuity of the limiting operator, it is more convenient for us to consider the topology given by the joint convergence of the first three powers of w_n. To be more precise, we require joint convergence of

$$\left(\int_{\mathbb{R}^d} f(x) w_n(x) dx, \int_{\mathbb{R}^d} f(x) w_n^2(x) dx, \int_{\mathbb{R}^d} f(x) w_n^3(x) dx \right)$$
$$\to \left(\int_{\mathbb{R}^d} f(x) w(x) dx, \int_{\mathbb{R}^d} f(x) w^2(x) dx, \int_{\mathbb{R}^d} f(x) w^3(x) dx \right).$$

We shall denote the topology generated by this convergence by τ. Note that the topology τ is *a priori* weaker than the topology of vague convergence.

The following discussion, whose main aim is to show that our choice does not alter compactness of the space, is a minor adaptation of Appendix A of *KALLENBERG* (2002, [26]). It is summarized by the following theorem.

Theorem C.1. *Consider the space $\mathcal{M}_\lambda(\mathbb{R}^d \times \mathcal{K})$ equipped with the topology τ. The following hold:*

i) The space $\mathcal{M}_\lambda$ equipped with the topology τ is Polish,

ii) A set $A \subset \mathcal{M}_\lambda$ is relatively compact with respect to τ if and only if for every $f \in C_c(\mathbb{R}^d)$

$$\sup_{\mu \in A} \int f \mathrm{d}\mu < \infty. \tag{C.1}$$

Therefore the space $\mathcal{M}_\lambda(\mathbb{R}^d \times \mathcal{K})$ is compact.

Proof. Consider a countable dense subset $\{f_i\}$ of $C_c(\mathbb{R}^d)$. For $\mu, \nu \in \mathcal{M}_\lambda(\mathbb{R}^d \times \mathcal{K})$ we define

$$d(\mu, \nu) = \sum_k 2^{-k} \left| \int f_k \mathrm{d}\mu - \int f_k \mathrm{d}\nu \right|. \tag{C.2}$$

49

Observe that $d(\mu, \nu)$ defines a metric on $\mathcal{M}_\lambda$. In particular, observe that since $w_n(x)$ and $w(x)$ are uniformly bounded and non-negative, the convergence of $\int w_n f(x) \mathrm{d}x$ implies convergence of $\int w_n^2 f(x) \mathrm{d}x$. Separability follows. Completeness will be clear once we prove ii).

Since integration against an element of $\mathcal{M}_\lambda$ is continuous for $f \in C_c$, the relative compactness of A implies (C.1). For the other direction, assume that the set $A \in \mathcal{M}_\lambda$ satisfies (C.1). Consider a family of compact subsets of $\mathbb{R}^d \times \mathcal{K}$ such that $K_n \subset int K_{n+1}$ and $\cup_n K_n = \mathbb{R}^d \times \mathcal{K}$. Consider a family of continuous cut-off functions $g_n \in C_c^\infty$, satisfying

$$g_n(x) = \begin{cases} g_n(x) = 1, x \in K_n \\ 0 \leq g_n(x) \leq 1, x \in K_{n+2} \\ g_n(x) = 0, x \neq K_{n+1} \end{cases}$$

We observe that for each n, the set $\{\int g_n \cdot \mu : \mu \in A\}$ is uniformly bounded and therefore, by Prochorov's Theorem, it is sequentially relatively compact. The compactness of $\mathcal{M}_\lambda$ follows by a diagonal argument.

$\square$

Remark C.2. *Our consideration can be extended to a wider class of spaces and topologies without any major changes. For example, one could take $\mathcal{K}$ to be any compact subset of $\mathbb{R}$ or require convergence of any higher power of the densities.*

D   Detailed description of simulations

D.1   A general description

We provide a more detailed description of the simulation.

Initialization   The simulation begins by choosing random seeds, separately for the times of reproduction/migration events and for the selection of individuals during those events. The user specifies the parameters of the simulation: the number of demes, their spatial structure (for the simulations reported in Section 7 this is always a one-dimensional torus), the number of individuals per deme, a correlation function for the environments, the selection coefficient, the rate of changes in the environment, the frequency of creating records and the time of the simulation. The rates of neutral and migration rates are determined by the number of individuals per deme, and the rate of selection events depends additionally on the selection coefficient.

The basic objects in the simulation are initialized. Each deme is assigned an identification number and a list of neighbours, determined by the selected spatial structure (for the one-dimensional torus, each deme has two neighbours). Each individual is randomly assigned a type according to a Bernoulli (1/2) distribution. Each individual is assigned an identification number, according to the deme in which it lives. This number allow us to track the ancestral origin of the population forward in time. The initial state of the environment in each deme is specified. A maximal time of the simulation is set.

The object that determines the times of the events, 'the Clock', is initialised. For each deme, for each type of event, the next time at which an event occurs is encoded. A separate variable with the time of the next environment change is also kept. The Clock is initialised with the exponential distribution with rates specified for each type of an event. The current time is set to 0.

Simulation   As long as the time of the simulation is smaller than the maximal time of the simulation the following is repeated.

The clock is asked to return the time, deme identification number α, and type κ of the next event (chosen to be the smallest number on the list of all times stored). This time is saved as the current time in the simulation. A random number is drawn from an exponential distribution with rate determined by the type κ of the event. This new time is added to the current time and passed to the Clock to specify the next time an event of type κ occurs in deme α.

The event of type κ is executed in deme with identification number α. With a given frequency (specified by the user with respect to the number of events that have taken place), a record of the current state of the population is created.

Events   There are four type of events: two types of reproduction events (neutral and selective), migration events and environmental events.

Whenever a reproduction event takes place in deme α, two individuals are chosen uniformly from deme α. One of the selected individuals is then randomly chosen to be a potential parent, using a Bernoulli (1/2) distribution. If the event is neutral, the potential parent is selected as the parent. If the event is selective, the types of the two selected individuals are checked. If they are the same, the potential parent is selected as the parent. If the types are different, the state of the environment is checked. Then the individual with type favoured by the environment is selected as the parent. The type and the ancestral origin of the parent are assigned to the second individual.

Whenever a migration event occurs in deme α, one of the neighbouring demes is randomly selected according to the uniform distribution. A single individual is chosen uniformly from each of deme α and the randomly selected neighbour and the type and ancestral origin of those individuals are swapped.

Whenever an environmental event occurs, a new state of the environment is randomly chosen. For the environmental regimes considered in Section 7, a single number (1 or -1) is drawn randomly from a Bernoulli (1/2) distribution. In the cases of neutral/constant selection, the environment is not changed. In the fluctuating cases, the environment is changed in all demes (consistent with the specified correlation function).

Recording   Two separate types of record are created, the 'proportion file' and the collection of 'ancestral origin files'. The number of ancestral origin files is equal to the number of demes, β.

The proportion file is a .txt file with $\beta + 3$ columns, separated by a single space. The first column contains the information of the time at which records are created. The second column contains the global proportion of type a at the given time. Columns $3, \ldots, (\beta + 3)$ record the proportion of type a in each deme.

Each 'ancestral origin' file is a .txt file with $\beta + 1$ columns, separated by a single space. File number γ contains information on the proportion of individuals in each deme with ancestral origin equal to γ. The first column contains the time of creation of the records. Columns $2, \ldots, (\beta + 1)$ record the proportion of individuals with ancestral origin equal to γ in each deme.

D.2   Implementation

The simulation was implemented in C++14. The Mersenne Twister pseudorandom number generator was used. Figures were created using R, (2015, [42]).

References

[1] S. Athreya and R. Tribe, *Uniqueness for a class of one-dimensional stochastic pde's using moment duality*, Ann. Probab. **28** (2000), no. 4, 1711–1734. MR 1813840

[2] N. H. Barton, *Genetic hitchhiking*, Phil. Trans. Roy. Soc. London B **355** (2000), 1553–1562.

[3] N. H. Barton, A. M. Etheridge, and A. Véber, *A new model for evolution in a spatial continuum*, Electron. J. Probab. **15** (2010), 162–216. MR 2594876

[4] ———, *Modelling evolution in a spatial continuum*, J. Stat. Mech. **P01002** (2013).

[5] A. O. Bergland, E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov, *Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in Drosophila*, PLoS Genetics **10** (2014), no. 11, e1004775.

[6] J. Bertoin and J.-F. Le Gall, *Stochastic flows associated to a coalescent process*, Prob. Theor. Rel. Fields **126** (2003), 261–288. MR 1990057

[7] M. Birkner, *Particle systems with locally dependent branching : long-time behaviour, genealogy and critical parameters*, Ph.D. thesis, Johann Wolfgang Goethe-Universität Frankfurt 2003.

[8] J. Blath, A. M. Etheridge, and M. Meredith, *Coexistence in locally regulated competing populations and survival of branching annihilating random walk*, Ann. Appl. Probab. **17** (2007), 1474–1507. MR 2358631

[9] I. Cvijovic, B. H. Good, E. R. Jerison, and M. M. Desai, *Fate of a mutation in a fluctuating environment*, Proc. Nat. Acad. Sci. U.S.A. **112** (2015), no. 36, E5021–E5028.

[10] L. F. Delph and J. K. Kelly, Tansley review: *on the importance of balancing selection in plants*, New Phytologist **201** (2014), 45–56.

[11] C. Doering, C. Mueller and P. Smereka, *Interacting particles, the stochastic Fisher-Kolmogorov-Petrovsky-Piscunov Equation, and duality*, Physica A **325** (2002), 243–259.

[12] P. J. Donnelly and T. G. Kurtz, *Particle representations for measure-valued population models*, Ann. Probab. **27** (1999), 166–205. MR 1681126

[13] R. Durrett and W.-T. Fan, *Genealogies in expanding populations*, Ann. Appl. Probab. **26** (2016), 3456–3490. MR 3582808

[14] A. M. Etheridge, *Drift, draft and structure: some mathematical models of evolution*, Banach Center Publ. **80** (2008), 121–144. MR 2433141

[15] A. M. Etheridge, A. Véber, and F. Yu, *Rescaling limits of the spatial Lambda-Fleming-Viot process with selection*, ARXIV1406.5884 (2019).

[16] A. Fijarczyk and W. Babik, *Detecting balancing selection in genomes: limits and prospects*, Molecular Ecology **24** (2015), 3529–3545.

[17] R. A. Fisher and E B Ford, *The spread of a gene in natural conditions in a colony of the moth Panaxia dominula l.*, Heredity **1** (1947), 143–174.

[18] J. H. Gillespie, *Genetic drift in an infinite population: the pseudohitchhiking model*, Genetics **155** (2000), 909–919.

[19] _____, *Is the population size of a species relevant to its evolution?*, Evolution **55** (2001), 2161–2169.

[20] _____, *Population genetics: a concise guide*, The Johns Hopkins University Press, 2nd Edition, 2004.

[21] Z. Gompert, *Bayesian inference of selection in a heterogeneous environment from genetic time-series data*, Molecular Ecology **25** (2016), 121–134.

[22] P. Hall, and C. C. Heyde, *Martingale limit theory and its application*, Academic press, (1980), MR 0624435

[23] O. Hallatschek and D. Nelson, *Gene surfing in expanding populations*, Theor. Pop. Biol. **73** (2008), 158–170.

[24] J. Hu, D. Nualart, and J. Song. A nonlinear stochastic heat equation: Hölder continuity and smoothness of the density of the solution. *Stochastic Processes and their Applications*, 123(3): 1083 – 1103, (2013). ISSN 0304-4149.

[25] M. Hutzenthaler, P. Pfaffelhuber, and C. Printz. *Stochastic averaging for multiscale Markov processes with an application to a Wright-Fisher model with fluctuating selection* ARXIV504.01508, (2018).

[26] O. Kallenberg. *Foundations of Modern Probability. Second Edition* Probability and Its Applications, Springer-Verlag, New York, 2002 MR 1876169

[27] S. M. Krone and C. Neuhauser, *Ancestral processes with selection*, Theor. Pop. Biol. **51** (1997), 210–237.

[28] T. G. Kurtz, *A limit theorem for perturbed operator semigroups with applications to random evolutions*, Journal of Functional Analysis **12** (1973), no. 1, 55–67. MR 0365224

[29] _____, *Averaging for martingale problems and stochastic approximation*, Applied Stochastic Analysis, vol. 177, Springer, Berlin, (1992). MR 1169928

[30] _____, *The Yamada-Watanabe-Engelbert theorem for general stochastic equations and inequalities* Electron. J. Probab. **12** (2007). MR 2336594

[31] _____, *Equivalence of stochastic equations and martingale problems*, Stochastic analysis 2010, Springer, (2010), pp. 113–130. MR 2789081

[32] _____, *Weak and strong solutions of general stochastic models.* Electron. Commun. Probab., **19** 16 pp., (2014). MR 3254737

[33] R. H. Liang, *Two continuum-sites stepping stone models in population genetics with delayed coalescence*, Ph D thesis, UC Berkeley, 2009.

[34] J. Maynard Smith and J. Haigh, *The hitch-hiking effect of a favourable gene*, Genet. Res. **23** (1974), pp. 23-35.

[35] L. Miller, *The effect of fluctuating selection in population genetics*, Transfer of Status dissertation, University of Oxford, 2012.

[36] C. Mueller and R. Tribe, *Stochastic p.d.e.'s arising from the long range contact and long range voter processes*, Prob. Th. Rel. Fields **102** (1995), 519–546. MR 1346264

[37] C. Neuhauser and S. M. Krone, *Genealogies of samples in models with selection*, Genetics **145** (1997), 519–534.

[38] L. Mytnik, E. Perkins and A. Sturm, *On pathwise uniqueness for stochastic heat equations with non-Lipschitz coefficients.* Ann. Probab. **34** (2006), no. 5, pp. 1910-1959. MR 2271487

[39] S. Novak and N. H. Barton, *When does frequency independent selection maintain genetic variation?*, Genetics **207** (2017), 653–668.

[40] R. B. O'Hara, *Comparing the effects of genetic drift and fluctuating selection on genotype frequency changes in the scarlet tiger moth*, Proc. Roy. Soc. London Ser. B **272** (2005), 211–217.

[41] V. V. Petrov, *On local limit theorems for sums of random variables*, Theor. Probab. Appl. **9** (1964), 312–320.

[42] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2015.

[43] T. Rippl. *Pathwise Uniqueness of the Stochastic Heat Equation with Holder continuous diffusion coefficient and colored noise.* PhD thesis, University of Göttingen, 2012.

[44] T. Rippl and A. Sturm. New results on pathwise uniqueness for the heat equation with colored noise. *Electron. J. Probab.*, 180 46 pp., (2013). MR 3101643

[45] T. Shiga, *Stepping stone models in population genetics and population dynamics*, Stochastic processes in physics and engineering (S Albeverio et al, ed.), D Reidel Publishing Company, 1988. MR 0948717

[46] J. E. Taylor, *The effect of fluctuating selection on the genealogy at a linked site*, Theor. Pop. Biol. **87** (2013), 34–50.

[47] R. Tribe, *Large time behavior of interface solutions to the heat equation with Fisher-Wright noise*, Prob. Th. Rel. Fields **102** (1995), 289–311. MR 1339735

[48] A. Véber and A. Wakolbinger, *The spatial Lambda-Fleming-Viot process: an event based construction and a lookdown representation*, Ann. Inst. H. Poincaré **51** (2015), 570–598. MR 3335017

[49] S. Wright, *Evolution and the genetics of populations, vol2: the theory of gene frequencies*, University of Chicago Press, 1969.