Some models from mathematical population genetics with fluctuations in the environment



Aleksander Klimek

Mathematical Institute

University of Oxford

This dissertation is submitted for the degree of

Doctor of Philosophy

St. Hilda's College

Trinity, 2018

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original. The results of Chapter 3 have been obtained in a joint effort with Niloy Biswas and Alison Etheridge. The results of Chapter 4 have been obtained in a joint effort with Jonathan Chetwynd-Diggle. This thesis is submitted to the University of Oxford for the degree Doctor of Philosophy, and has not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

Aleksander Klimek

Trinity, 2018

Acknowledgements

First and foremost, I would like to thank my supervisor Alison Etheridge, for her support, guidance and patience. Secondly, I would like to thank the OxPDE group and especially the PDE CDT directors, Gui-Qiang Chen, John Ball and Endre Süli, without them this adventure (admittedly on the fringes of their interests) would not be possible. I would also like to thank Ben Hambly and Amandine Véber for kindly agreeing to examine my thesis; Tom Kurtz for many helpful discussions; and Jono Chetwynd-Diggle for our fruitful collaboration.

I am also grateful to my family and friends, for their constant encouragement and understanding. I would like to thank Bogdan for all our chats and a marvellous table football rivalry. My time in Oxford would not be the same without Franz, Panu, Nick, Collin, Guangyu, Hannah, Andrew and Jen, thank you for sharing your time with me.

Finally I would like to thank my parents and grandparents for their patience, support and advice, and Joanna for everything.

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 partial differential equation 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 Hallatschek and Nelson (2008) and Durrett and Fan (2016), in numerical experiments with a closely related model based on the classical Moran model. We also investigate the behaviour of an establishing mutation which is subject to rapidly fluctuating selection under the Lambda-Fleming-Viot model and show that under a suitable scaling it converges to the Feller diffusion in a random

environment. We then extend to a population that is distributed across a spatial continuum. In this setting the scaling limit is the SuperBrownian motion in a random environment. The scaling results for the behaviour of the rare allele are achieved via particle representations which belong to the family of 'lookdown constructions'.

Table of contents

1 Introduction **1**

1.1 Biological background 4

1.2 Description of the models 7

1.3 Statement of main results 11

1.4 Numerical results 14

1.5 Markov Mapping Theorem 19

1.6 Structure of the thesis 22

2 Mathematical Prerequisites **25**

2.1 Some facts from functional analysis 29

2.1.1 Trace class operators 30

2.1.2 Hilbert-Schmidt operators 31

2.1.3 Gaussian measures and processes on Hilbert spaces 32

2.2 Stochastic integral 34

2.2.1 Integrators 34

2.2.1.1 Q-Wiener measure 35

2.2.1.2 Cylindrical Brownian motion 36

2.2.1.3 Worthy martingale measures 37

2.2.1.4 Gaussian coloured noise as a Q-Wiener measure 43

2.2.2 Stochastic integral 44

2.2.2.1 Integral with respect to Q-cylindrical Wiener measure . 44

2.2.2.2 Integral with respect to a worthy martingale measure . 47

2.2.3 Further extensions of the stochastic integral 49

2.3 Yamada-Watanabe Theorem 51
2.3.1 Yamada-Watanabe theorem for SDEs 51
2.3.1.1 Different notions of solution for SDEs 51
2.3.1.2 Weak solutions and martingale problems 53
2.3.1.3 Notions of uniqueness of solution for SDEs 55
2.3.2 Illustrative examples 56
2.3.3 Classical Yamada-Watanabe Theorem 57
2.3.3.1 An overview of Yamada-Watanabe results for more general SDEs 58
2.4 Solutions of general stochastic models 59
2.4.1 Framework and General Yamada-Watanabe Theorem 60
2.4.2 Notions of solutions for SPDE 63
2.4.3 General Yamada-Watanabe Theorem and SPDEs 67
2.5 Stochastic heat equation 69
2.6 Equivalence of the martingale problem and SPDE 71
2.7 Poisson random measures 76

3 Spatial Lambda-Fleming-Viot Model in a fluctuating environment 79
3.1 The non-spatial case 79
3.1.1 The (non-spatial) model 79
3.1.2 Scaling limits 81
3.2 Definition and scaling of the Spatial Lambda-Fleming-Viot process with fluctuating selection 85
3.2.1 Spatial Lambda-Fleming-Viot process with fluctuating selection . 85
3.2.2 Scaling the SLFVFS 91
3.3 Duality 95
3.4 Tracer dynamics 98
3.5 Proof of Theorem 3.10 102
3.5.1 Identifying the limit 104
3.5.2 Uniqueness of solutions for limiting equation 115
3.5.3 Proof of Proposition 3.20 116

4 Lookdown constructions **121**

4.1 Introduction . . . 121

4.2 SuperBrownian motion in a random environment . . . 126

4.2.1 Definitions . . . 127

4.2.2 Lookdown construction for BBMRE and SBMRE . . . 130

4.3 Spatial Lambda-Fleming-Viot model . . . 142

4.4 Scaling limits of the LFV - dynamics of the rare type . . . 148

4.4.1 Scaling . . . 152

4.4.2 Statement of main results . . . 153

4.4.2.1 Projected model . . . 156

4.4.3 Convergence of the non-spatial model . . . 157

4.4.3.1 Intensity estimate . . . 161

4.4.3.2 Neutral model - proof of Theorem 4.20 . . . 164

4.4.3.3 Selective model - proof of Theorem 4.21 . . . 177

4.4.3.4 Model with selection in fluctuating environment - proof of Theorem 4.18 . . . 182

4.5 Scaling limits of the SLFV - dynamics of the rare type . . . 183

4.5.1 Scaling and statement of main results . . . 185

4.5.2 Intensity estimate . . . 188

4.5.3 Sketch of the proof of Theorem 4.43 . . . 192

Bibliography **197**

Appendix A Markov Mapping Theorem **203**

Appendix B Numerical Simulations **207**

B.1 Detailed description of simulations . . . 207

B.1.1 A general description . . . 207

B.2 Implementation . . . 210

Chapter 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 thesis 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), *"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, 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 nontrivial frequencies for long periods of time. If the population has no spatial structure, then over large timescales (thousands of generations) the frequency

of a-alleles can be modelled (see, i.e. Kimura (1954), Gillespie (1977), Huerta-Sanchez et al. (2008)) by a stochastic differential equation:

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

where B^1 and B^2 are independent Brownian motions, the first (as we shall explain in Section 3.1) 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 3.1). For a population distributed across a one-dimensional spatial continuum, one can write down an analogous Stochastic Partial Differential Equation (SPDE):

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

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 (see discussion at the beginning of Chapter 2), and we must find a different approach.

The difficulties with modeling genetic drift in populations evolving in higher dimensional spatial continua, often referred to as 'the pain in the torus', are well known; see Barton et al. (2013) for a review. They can be overcome using the spatial Lambda-Fleming-Viot process, introduced in Etheridge (2008) and rigorously constructed in Barton et al. (2010), and here we adapt that approach 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) 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 SPDE (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 an SPDE 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); Neuhauser and Krone (1997), 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 3.3 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) and, more recently, Durrett and Fan (2016) 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. This idea is exploited in our numerical experiments of Section 1.4.

In a different line of investigation we address another question of some biological interest: how does the frequency of a rare mutation evolve in a spatially distributed population? What if the direction of selection fluctuates in time?

Consider for a moment a model without selection. Suppose that the proportion of the population carrying allele a is small. In the absence of spatial structure, the absolute number of such individuals evolves approximately according to a branching process,

which, under appropriate scaling, converges to the Feller diffusion. We may ask whether a similar phenomenon occurs in the presence of selection, especially when its direction fluctuates in time. We expect that in the latter case the evolution follows a branching process in a random environment, and, under suitable scaling, converges to the Feller diffusion in a random environment. We investigate that question in Chapter 4.

It is natural to try to establish an analogous result for a spatially distributed population. This question has been already addressed for the neutral spatial Lambda-Fleming-Viot model. In Chetwynd-Diggle and Etheridge (2018) it is shown that the suitably scaled model converges to a SuperBrownian motion. Their proof is based on a duality argument. Since such a technique is not available to us, we use a different approach, which belongs to the family of lookdown constructions, discussed in Chapter 4.

We show that a suitably rescaled version of a Spatial Lambda-Fleming-Viot with selection in a fluctuating environment converges to a SuperBrownian motion in a random environment. The original construction and study of this process is based on approximate duality (see Mytnik (1996a)). We provide a new construction based on the lookdown approach in Section 4.2. The scaling result can be found in Section 4.5.

1.1 Biological background

Suppose that a gene occurs in just two forms that, because of environmental fluctuations, 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) 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) 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) 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) 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. This conclusion is consistent with the 'Court Jester hypothesis' , that long term improvements in fitness may not occur, because populations have to 'run to stand still' in order to keep pace with changes in the environment, see Barnosky (2001), Benton (2009).

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 selecion 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). Of course, correspondingly, a neutral allele associated with an unfavoured type will decrease in frequency. Gillespie (2000, 2001) 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) 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). 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) shows that the effects on the genealogy at a linked neutral locus will differ. Fijarczyk and Babik (2015) 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 et al. (2014) 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) 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 et al. (2015) 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.

1.2 Description of the models

We begin with a non-spatial version of model which is most important in the thesis. The following version of the Lambda-Fleming-Viot model is a simple extension of the model introduced in Donnelly and Kurtz (1999a), Bertoin and Le Gall (2003).

Definition 1.1 (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* $\mathrm{d}t \otimes \nu(\mathrm{d}u) \otimes \sigma(\mathrm{d}\mathbf{s})$*, 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*

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

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

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

This model is discussed in detail in Section 3.1. The next model is a spatial counterpart of the Lambda-Fleming-Viot model with fluctuating selection. It is an extension of the Spatial Lambda-Fleming-Viot model with selection introduced and studied in Etheridge et al. (2014).

Definition 1.2 (Spatial Lambda-Fleming-Viot process with fluctuating selection)**.** *Let* μ *be a* σ*-finite 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*

$$\int_{(0,\infty)} r^d \int_{[0,1]} u\, \nu_r(\mathrm{d}u)\mu(\mathrm{d}r) < \infty. \tag{1.3}$$

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* $q(x,y)$ *be a covariance function which is an element of* $C_b\Big(\mathbb{R}^d \times \mathbb{R}^d\Big)$ *(space of bounded continous functions on* $\mathbb{R}^d \times \mathbb{R}^d$*) and let* $\{\xi^{(m)}(\cdot)\}_{m\geq 0}$ *be a family of identically distributed random fields such that*

$$\mathbb{P}\Big[\xi^{(m)}(x) = -1\Big] = \frac{1}{2} = \mathbb{P}\Big[\xi^{(m)}(x) = +1\Big],$$
$$\mathbb{E}\Big[\xi^{(m)}(x)\xi^{(m)}(y)\Big] = q(x,y).$$

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 $\xi(t,\cdot)$ *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 process* M_t, *taking values in* $\mathcal{M}_\lambda$*(the space of measures on* $\mathbb{R}^d \times \{a, A\}$ *with first component being equal to Lebesgue measure), 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 distribution*

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

(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 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*

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

(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 $y \in B(x,r)$,*

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

This model is discussed in detail in Section 3.2.

The next model that plays an important role is SuperBrownian motion in a random environment, introduced and studied in Mytnik (1996b). In Section 4.2 we provide a new description of this model in terms of a lookdown construction. In Section 4.5 we show that this model captures the behaviour of an establishing mutation which is subject to rapidly fluctuating selection under suitably rescaled Spatial Lambda-Fleming-Viot model with fluctuating selection. Let us introduce some notation. For a measure X,

defined on $\mathbb{R}^d$, we define $X(f)$ as

$$X(f) = \int_{\mathbb{R}^d} f(x)X(\mathrm{d}x).$$

Definition 1.3 (SuperBrownian motion in a random environment (SBMRE))**.** *Let $q(x,y) \in C_0(\mathbb{R}^d \times \mathbb{R}^d)$ be a symmetric function. The superBrownian motion in a random environment is the (unique) process, taking values in $\mathcal{M}_f(\mathbb{R}^d)$ (that is, the space of finite measures on $\mathbb{R}^d$), for which, for all $\phi \in \mathcal{D}(\Delta)$,*

$$X_t(\phi) = X_t(\phi) - X_0(\phi) - \int_0^t \frac{1}{2} X_s(\Delta\phi)\mathrm{d}s$$

is a square-integrable martingale with quadratic variation given by

$$\langle X(\phi)\rangle_t = \int_0^t X_s(\phi^2)\mathrm{d}s + \int_0^t \int_{\mathbb{R}^d \times \mathbb{R}^d} q(x,y)\phi(x)\phi(y)X_s(\mathrm{d}x)X_s(\mathrm{d}y)\mathrm{d}s.$$

1.3 Statement of main results

In this section we state the main results of this thesis. We begin with a result which describes the diffusion approximation of the Spatial Lambda Fleming Viot model with fluctuating selection.

Definition 1.2 is rather general. We include it to underline the possibility of extending our results. In the interest of avoiding overly complicated notation, from now on we shall specialise to fix the radius and impact of reproduction events.

Assumption 1.4. *From now on, fix $R \in (0,\infty)$ and $\bar{u} \in (0,1)$ and take*

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

Just as in Etheridge et al. (2014), we shall prove convergence, not of the sequence of densities of the process, but of a sequence of local averages. We require some notation. Let R be the fixed radius of events. Set

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

and define the sequence of rescaled processes

$$w_n(t,x) = w(nt, n^{\frac{1}{3}}x), \quad \overline{w}_n(t,x) = \fint_{B_n(x)} w_n(t,y)\mathrm{d}y, \tag{1.4}$$

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

Theorem 1.5. *Write $(\overline{M}^n_t)_{t\geq 0}$ for the measure-valued process with density $(\overline{w}_n(t))_{t\geq 0}$. Suppose that $(M^n_0)_{n\geq 1}$ converges weakly in $\mathcal{M}_\lambda$ to the measure with density $\lim_{n\to\infty}\overline{w}_n(0,x)$. Further, fix $\alpha \in (0,1/6)$, set $s_n = \mathbf{s}n^\alpha/n^{2/3}$, $u_n = \bar{u}/n^{1/3}$ 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 $q(x,y)$ that determines the environment satisfies*

$$|q(x,x) - q(x,y)| \leq C|x-y| \qquad \text{for all } x,y \in \mathbb{R}^d. \tag{1.5}$$

Then the sequence $(\overline{M}^n_\cdot)_{n\geq 1}$ 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 $(M^\infty_t)_{\geq 0}$, writing w^∞ for a representative of the density of M^∞,

1. *for dimension $d=1$, w^∞ is the continous in space process for which, for every $F \in C^\infty_c(\mathbb{R})$ (continous, compactly supported function) and for every $f \in C^2_c(\mathbb{R})$ (continous, twice differentiable function with compact support),*

$$\begin{aligned}
&F(\langle w^\infty(t), f\rangle) - F(\langle w^\infty(0), f\rangle) \\
&\quad - \int_0^t F'(\langle w^\infty(s), f\rangle)\Big\{\Big\langle w, \frac{\bar{u}\Gamma_R}{2}\Delta f\Big\rangle \\
&\qquad\qquad + V_R^2\bar{u}^2\mathbf{s}^2\langle w^\infty(s)(1-w^\infty(s))(1-2w^\infty(s)), f\rangle\Big\}\mathrm{d}s \\
&\quad - \int_0^t F''(\langle w^\infty(s), f\rangle)\Big\{ \\
&\qquad\qquad \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Big[V_R^2\bar{u}^2\mathbf{s}^2 q(x,y) w^\infty(s,x)(1-w^\infty(s,x)) \\
&\qquad\qquad\qquad \times w^\infty(s,y)(1-w^\infty(s,y))f(x)f(y)\Big]\mathrm{d}x\mathrm{d}y \\
&\qquad\qquad\qquad + \langle 2\bar{u}^2R^2 w^\infty(s)(1-w^\infty(s)), f^2\rangle\Big\}\mathrm{d}s
\end{aligned} \tag{1.6}$$

is a martingale;

2. for dimension $d \geq 2$, for every $F \in C_c^\infty(\mathbb{R})$ and for every $f \in C_c^2\Big(\mathbb{R}^d\Big)$,

$$\begin{aligned} &F(\langle w^\infty(t), f\rangle) - F(\langle w^\infty(0), f\rangle) \\ &\quad - \int_0^t F'(\langle w^\infty(s), f\rangle)\Big\{\Big\langle w, \frac{\bar{u}\Gamma_R}{2}\Delta f\Big\rangle \\ &\qquad\qquad + V_R^2\bar{u}^2\mathbf{s}^2\langle w^\infty(s)(1-w^\infty(s))(1-2w^\infty(s)), f\rangle\Big\}\mathrm{d}s \\ &\quad - \int_0^t F''(\langle w^\infty(s), f\rangle)\Big\{ \\ &\qquad \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Big[V_R^2\bar{u}^2\mathbf{s}^2 q(x,y)w^\infty(s,x)(1-w^\infty(s,x)) \\ &\qquad\qquad \times w^\infty(s,y)(1-w^\infty(s,y))f(x)f(y)\Big]\mathrm{d}x\mathrm{d}y\Big\}\mathrm{d}s \quad (1.7) \end{aligned}$$

is a 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 (3.37)*.*

We also show that every solution to the martingale problem in $d \geq 2$ is a weak solution to the corresponding SPDE:

$$\begin{aligned} \mathrm{d}w^\infty = &\Big(\frac{\bar{u}\Gamma_R}{2}\Delta w^\infty + \bar{u}^2V_R^2\mathbf{s}^2w^\infty(1-w^\infty)(1-2w^\infty)\Big)\mathrm{d}t \\ &+ \sqrt{2}\bar{u}V_R\mathbf{s}w^\infty(1-w^\infty)W(\mathrm{d}t,\mathrm{d}x), \quad (1.8) \end{aligned}$$

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

$$\langle W(\phi)\rangle_t = t\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} q(x,y)\phi(x)\phi(y)\mathrm{d}x\mathrm{d}y. \quad (1.9)$$

The corresponding equation for dimension $d = 1$ is

$$\begin{aligned} dw^\infty = &\Big(\frac{\bar{u}\Gamma_R}{2}\Delta w^\infty + 4R^2\bar{u}^2\mathbf{s}^2w^\infty(1-w^\infty)(1-2w^\infty)\Big)dt \\ &+ \sqrt{2}2R\bar{u}\mathbf{s}w^\infty(1-w^\infty)W(\mathrm{d}t,\mathrm{d}x) + 2R\bar{u}\sqrt{w^\infty(1-w^\infty)}\mathcal{W}(\mathrm{d}t,\mathrm{d}x), \quad (1.10) \end{aligned}$$

where W is white in time and coloured in space as above, and $\mathcal{W}$ is a space-time white noise. However, our proof does not guarantee that the solutions to martingale problem are equivalent to solutions to the SPDE in that case.

The proof of this Theorem is the content of Section 3.5. An analogous result for the non-spatial version of the model can be found in Section 3.1. In Section 4.4 we show the following:

Theorem 1.6. *Let $X^N(t)$ denote the total intensity of individuals of rare type at time t. Suppose that X_0^N converges to X_0. Then, under a scaling specified in Section 4.4, the sequence $X_N(t)$ converges weakly to the Feller diffusion in a random environment.*

A spatial analogue of this Theorem is the subject of Section 4.5. The limiting object is a superBrownian motion in a random environment with a drift.

1.4 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 (Moran (1958)) to incorporate spatial structure and fluctuating selection.

Definition 1.7 (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* $\mathbb{E}[\zeta(t,x)\zeta(t,y)] = q(x,y)$ *for a correlation function* q.

Remark 1.8. *As our lattice contains only finetly many points any correlation function can be chosen.*

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, that is, the environment fluctuates in time and the direction of selection is always different in demes $0-50$ and $51-100$;

2. The direction of selection fluctuates in time but is the same in every deme;

3. Constant selection (the environment is fixed), with direction of selection in demes $0-50$ opposite to 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 precise description of the code used for simulations can be found in Appendix B.

As a first comparison, Figure 1.1 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.

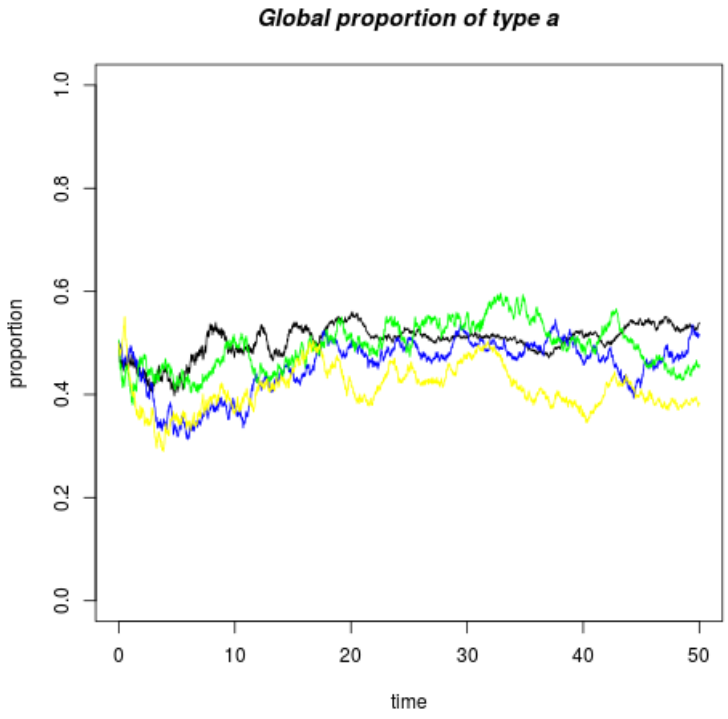


Fig. 1.1 Global proportions of type a individuals; one realisation, four different scenarios: blue anticorrelated environments, green correlated environments, black constant selection, and yellow neutral. See the main text for a full explanation.

In Figure 1.2, 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 left 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.

The overall proportion 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

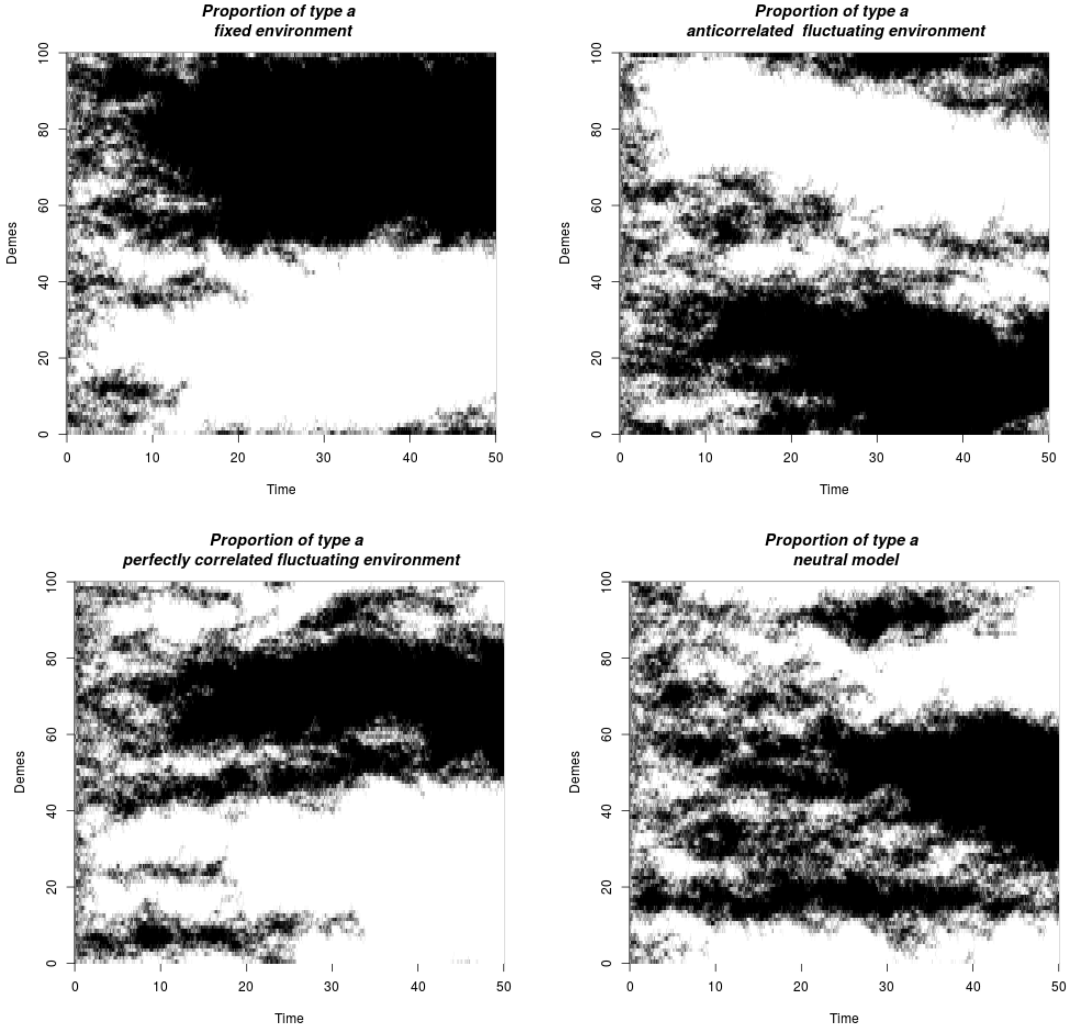


Fig. 1.2 Map of the proportion of type a individuals in each deme; the darker the colour the higher the proportion of type a. Top left, constant selection; top right, fluctuating selection with demes $0-50$ and $51-100$ perfectly anticorrelated; bottom left, fluctuating selection with all demes perfectly correlated; bottom right, the neutral case.

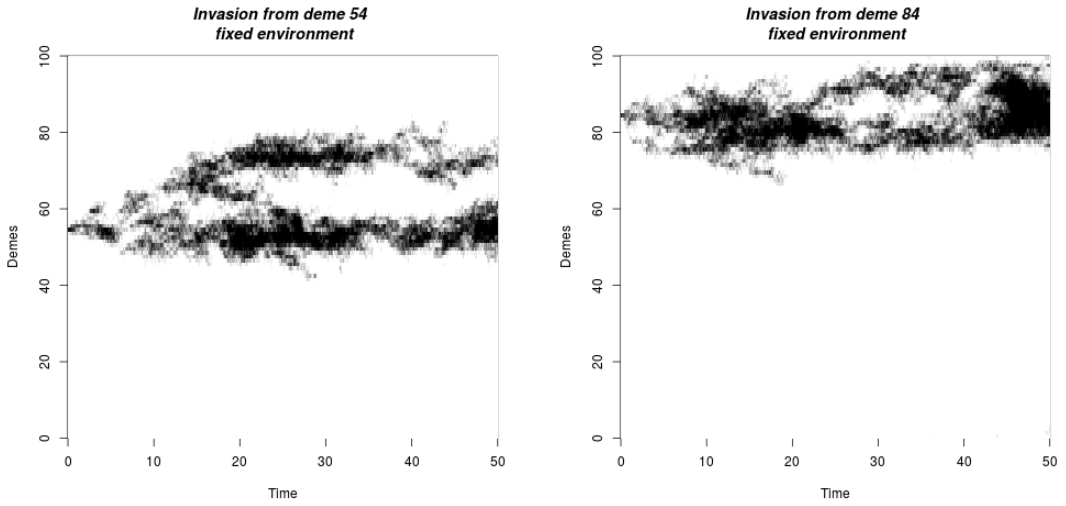


Fig. 1.3 Tracer map of individuals descended from ancestors located in particular demes at time zero; the darker the colour, the higher the proportion of such individuals. Fixed environment.

'tracers', further explored in the context of Spatial Lambda-Fleming-Viot process in Chapter 3, Section 3.4. In Figure 1.3 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, 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 (right hand panel).

In left panel of Figure 1.4, 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 1.4 can be contrasted with the left hand panel of Figure 1.5. 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

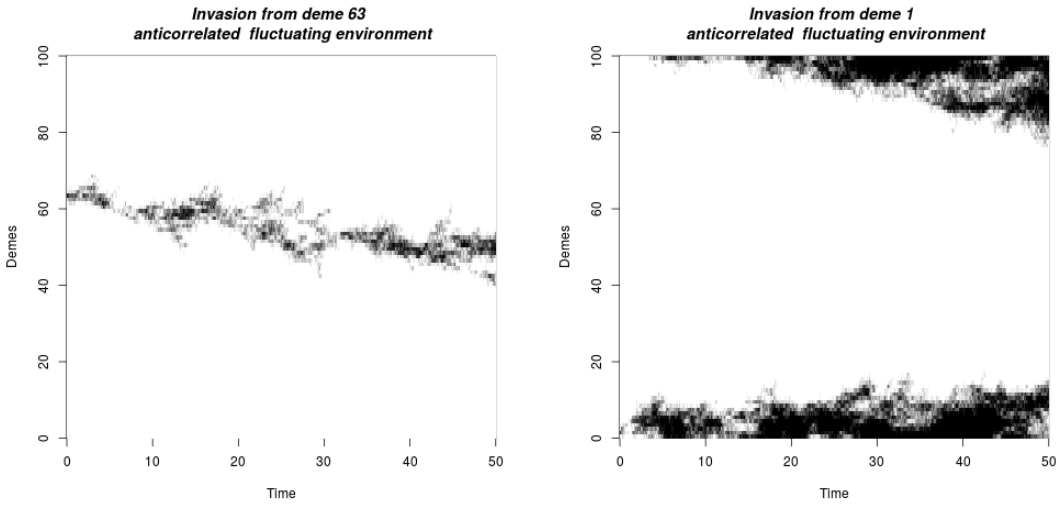


Fig. 1.4 Tracer map of individuals descended from ancestors located in particular demes at time zero; the darker the colour, the higher the proportion of such individuals. The environment is fluctuating, with its value in demes $0-50$ perfectly anticorrelated with that in demes $51-100$.

Figure 1.5 shows the 'thinning' of the family resulting from the periods of time when it is not favoured.

Finally, Figure 1.6, 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.

1.5 Markov Mapping Theorem

A crucial technical tool used in several points of the thesis is the Markov Mapping Theorem, introduced in Kurtz (1998) (see Theorem 2.7 for the original statement). The following version of the theorem is used extensively in this thesis.

Theorem 1.9 (Kurtz and Rodrigues (2011), Theorem A.15)**.** *Let (E,d) and (E_0,d_0) be a pair of complete, separable metric spaces. Let $A \subset \overline{C}(E) \times C(E)$ and let ψ be a continuous function taking values in $\mathbb{R}$ such that $\psi \geq 1$. Suppose that for each $f \in \mathcal{D}(A)$*

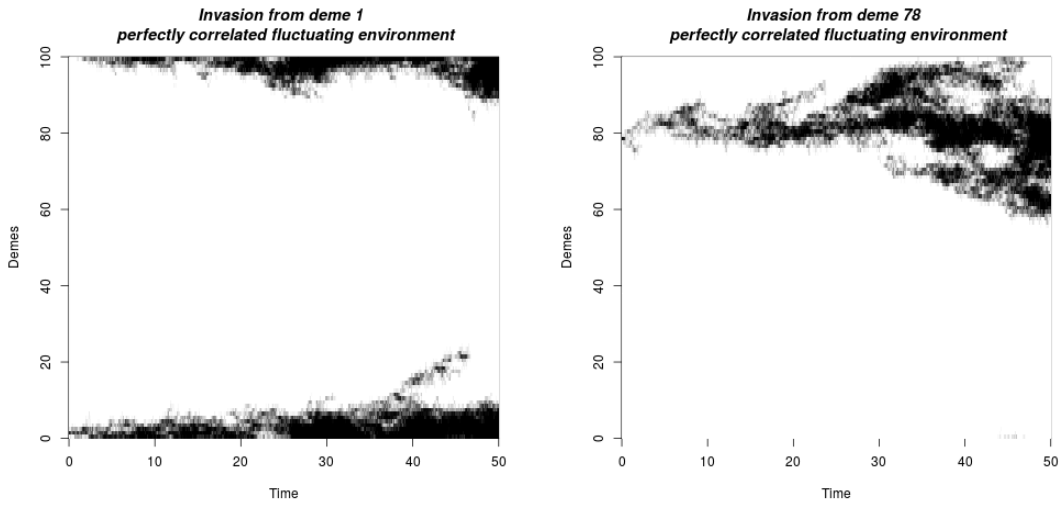


Fig. 1.5 Tracer map of individuals descended from ancestors located in particular demes at time zero; the darker the colour, the higher the proportion of such individuals. The environment is fluctuating, with its value in demes $0-50$ perfectly correlated with that in demes $51-100$.

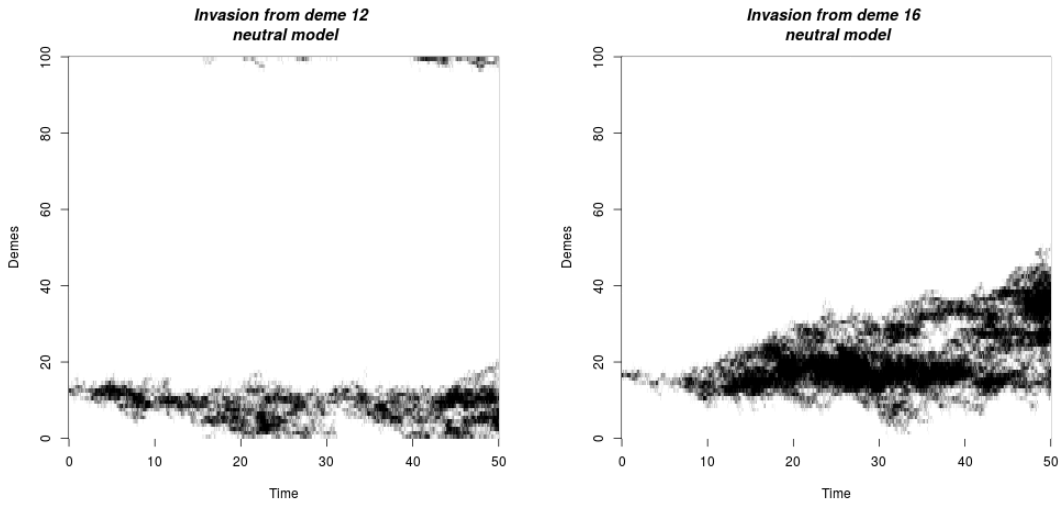


Fig. 1.6 Tracer map of individuals descended from ancestors located in particular demes at time zero; the darker the colour, the higher the proportion of such individuals. Neutral model.

there exists a constant $c_f > 0$ such that

$$|Af(x)| \leq c_f \psi(x).$$

Let A_0 be defined as

$$A_0 f(x) = \frac{Af(x)}{\psi(x)}.$$

Suppose that A_0 is a graph-separable pre-generator (see Definition A.2) and suppose that $\mathcal{D}(A) = \mathcal{D}(A_0)$ is closed under multiplication and separating. Let $\gamma : E \to E_0$ be Borel measurable, and let α be a transition function from E_0 into E (that is, $E_0 \ni y \mapsto \alpha(y, \cdot) \in \mathcal{P}(E)$ is Borel measurable) satisfying $\alpha(y, \gamma^{-1}(y)) = 1$. Assume that for each $y \in E_0$

$$\widetilde{\psi}(y) = \int_E \psi(y, z)\alpha(y, \mathrm{d}z) < \infty,$$

and define

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

Let $\mu_0 \in \mathcal{P}(E_0)$, and define $\nu_0(y) = \int_{\mathbb{R}} \alpha(y, \cdot)\mu_0(\mathrm{d}y)$.

1. *Let $\tilde{Y}$ be the solution of a martingale problem for (C, μ_0). Assume that it satisfies the moment condition*

$$\int_0^t \mathbb{E}\left[\widetilde{\psi}(\tilde{Y}(s))\mathrm{d}s\right] < \infty \quad \forall t \geq 0. \tag{1.11}$$

   Then there exists a solution X of the martingale problem (A, ν_0) such that $\widetilde{Y}$ has the same distribution on the space of Borel measurable functions from $\mathbb{R}$ to E as $Y = \gamma \circ X$.

2. *If, in addition, uniqueness holds for the martingale problem for (A, ν_0), then uniqueness holds for the martingale problem (C, μ_0).*

The original proof of the theorem was inspired by the proofs of generalisations of Burke's Output Theorem appearing in Kliemann et al. (1990) and the proof of equivalence of martingale problems for the Moran model and its lookdown representation in Donnelly and Kurtz (1996). We briefly discuss the proof of the latter which uses an application of the Markov Mapping Theorem at the begining of Chapter 4.

Since then the Markov Mapping Theorem has been a useful tool in mathematical population genetics (Etheridge and Kurtz (2018), Kurtz and Rodrigues (2011)), mathematical biology (Gupta (2012)), mathematical finance (Stockbridge (2002)) and analysis of infinite dimensional stochastic differential equations, Kurtz (2010).

The main power of the Markov Mapping Theorem comes in simplifications of proofs of equivalence of seemingly different martingale problems. The main source of the power is in exploitation of properties of exchangeable process and conditionally Poisson systems.

In this thesis we use the Markov Mapping Theorem to extend the proof of equivalence of martingale problems for infinite dimensional stochastic differential equations and Itô diffusions to SPDEs. The Markov Mapping Theorem is also the fundamental tool in Chapter 4, since it allows us to show that the lookdown representations we discuss do describe the processes we intended.

1.6 Structure of the thesis

The rest of the thesis is laid out as follows. In Chapter 2 we discuss some of the background material on SPDE. In Section 2.1 we recall some basic facts from Functional Analysis. In Section 2.2 we sketch the construction of infinite dimensional noises and integrals with respect to them. Section 2.3 is devoted to the classical Yamada-Watanbe Theorem and Section 2.4 to a generalisation of this theorem due to Kurtz (2007, 2014). In particular we prove a lemma which allows us to apply this theory to SPDEs. In Section 2.5 we state useful results on the Stochastic Heat Equation. This section contains an extension to SPDE of a result due to Kurtz (2010), which shows that weak solutions to Stochastic Differential Equations are equivalent to solutions to martingale problems for corresponding Itô diffusions.

Chapter 3 is devoted to the Spatial Lambda-Fleming-Viot process with fluctuating selection. We begin by investigating the model in the absence of spatial structure in Section 3.1. Section 3.2 describes the model, scaling and main results. Section 3.3 is devoted to the rather limited cases in which a duality argument is available. Section 3.4 contains a description of the 'tracer dynamics' (dynamics of a marked subpopulation). Section 3.5 is devoted to the proof of the main result.

Chapter 4 is devoted to lookdown constructions. We begin by explaining the main ideas for the Moran model and for a pure death process. In Section 4.2 we present a new lookdown construction of SuperBrownian motion in a random environment. In Section 4.3 we recall two constructions of the neutral Spatial Lambda-Fleming-Viot model. In Section 4.5 we present a scaling result for the Spatial Lambda-Fleming-Viot model with fluctuating selection, using its lookdown representation. The limiting object is SuperBrownian motion in a random environment.

A detailed descrition of the numerical simulations and the code can be found in Appendix B. Several versions of the Markov Mapping Theorem used in the thesis are stated in Appendix A.

Chapter 2

Mathematical Prerequisites

In this chapter we discuss some background material on stochastic processes. First, we discuss stochastic differential equations in infinite dimensional spaces. In particular, we include the material which allows us to make sense of a Stochastic Heat Equation (SHE) driven by white noise and an SHE driven by noise which is white in time and coloured in space. We also describe some basic properties of Poissson random measures, which are later used in Chapter 4.

We present two basic approaches to understanding stochastic integrals - one based on martingale measures, discussed extensively by Walsh (1986), and one based on the interpretation of an SPDE as an infinite dimensional system of SDEs, in the spirit of Da Prato and Zabczyk (2008). Dependent on the point of view, it is convenient to write the SHE in the form

$$\mathrm{d}u = \Delta u + b(t,x,u)\mathrm{d}t + \sigma(t,x,u)W(\mathrm{d}t,\mathrm{d}x), \tag{2.1}$$

when considering the martingale measure solutions, and

$$\mathrm{d}X(t) = [AX(t) + B(t,X(t))]\mathrm{d}t + \sigma(t,X(t))\mathrm{d}W(t), \tag{2.2}$$

where A is the generator of a semigroup (for most of our applications it will be a Lapalacian), B is a measurable map between Hilbert spaces and σ is a bounded map on Hilbert space, in the context of Hilbert spaces. The main difference between the two

is the fact that for (2.1), we interpret the solution as a random field, whereas for (2.2), we interpret the solution as a random element of a Hilbert space (or, in a more general setup, a Banach space).

The main difficulty, in comparison to the deterministic versions of equations (2.1), (2.2), comes from the random perturbation W. In this thesis we mostly remain in the realm where the random term is a centred Gaussian random variable. We distinguish between two basic cases. The first one, which represents the situation where the random perturbation is completely uncorrelated in time and space, is the so-called white noise case. The white noise can be (not entirely rigorously) described as a centred Gaussian random variable with correlation function

$$\mathbb{E}[W(t,x)W(s,x)] = (t \wedge s)\delta_x\delta_y,$$

where δ_z denotes the Dirac delta (a point measure) at point z and $(t \wedge s)$ is the minimum of t and s. The other class of noises we consider are the coloured noises - these are still completely uncorrelated (white) in time, but they have a correlation structure in space. We may write their correlation function (again, not entirely rigorously) as

$$\mathbb{E}[W(t,x)W(s,x)] = (t \wedge s)q(x,y),$$

where $q(x,y) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a sufficiently regular function (for most of our purposes, a bounded continuous function, or a square integrable function). The function $q(x,y)$ is often referred to as the covariance function.

To illustrate some of the problems caused by the random perturbation let us focus on the simplest case - the linear heat equation with additive noise on $\mathbb{R}^d$, that is

$$\mathrm{d}u = \Delta u + W(\mathrm{d}t, \mathrm{d}x). \tag{2.3}$$

Our analysis is not entirely rigorous. Assume that the initial data is identically equal to 0. If that is the case, the solution to SHE (2.3) is given by convolution with a

fundamental solution. To be more explicit, if we define $p_t(x)$ as

$$p_t(x) = \left(\frac{1}{4\pi t}\right)^{\frac{d}{2}} \exp\left(-\frac{|x|^2}{4}\right) \tag{2.4}$$

the solution to equation (2.3) is given by

$$u(t,x) = \int_0^t \int_{\mathbb{R}^d} p_{t-s}(x-y) W(\mathrm{d}s, \mathrm{d}y). \tag{2.5}$$

The first difficulty we face is giving meaning to the integral appearing in (2.5). For the next few lines let us assume that we have succeeded. Then we would expect that $u(t,x)$ given by (2.5) is a mean zero Gaussian process. Let us evaluate the variance of this process. If W is white noise, the covariance is given by

$$\begin{aligned}
\mathbb{E}[u(t,x)u(s,y)] &= \mathbb{E}\left[\int_0^t \int_0^s \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} p_{t-t'}(x-x') p_{s-s'}(y-y') W(\mathrm{d}s', \mathrm{d}y') W(\mathrm{d}t', \mathrm{d}x')\right] \\
&= \left(\frac{1}{4\pi}\right)^d \int_0^{(s\wedge t)} \int_{\mathbb{R}^d} \left(\frac{1}{|s-s'||t-s'|}\right)^{\frac{d}{2}} \exp\left(-\frac{|x-y|^2}{4(s-s')}\right) \exp\left(-\frac{|y|^2}{4(t-s')}\right) \mathrm{d}y \mathrm{d}s' \\
&= 2^{-(d+1)} \int_{|s-t|}^{s+t} (s')^{-\frac{d}{2}} \exp\left(\frac{|x-y|^2}{t-s'}\right),
\end{aligned}$$

which is finite if and only if $d = 1$. So the solution of (2.3) has finite variance only in dimension 1. We could still try to give a meaning to the solution in the sense of distributions, but since we do not expect that the regularity of the solution improves when the equation becomes more complicated (e.g. b and σ are functions, possibly nonlinear, of the solutions), if we were to study the nonlinear versions of SHE (2.1) in dimensions $d > 1$, we would have to give a meaning to a nonlinear function of a distribution. For this reason the analysis of SHE (2.1) is usually restricted to dimension $d = 1$.

An analogous computation could have been performed for noise which is coloured in space and white in time. To ensure the finiteness of the variance, we need to impose some conditions on the covariance kernel $q(x,y)$. It turns out that it is enough to assume that the $q(x,y)$ is square integrable or that it satisfies the *Dalang condition*, see Subsection 2.2.3 for some details.

Another set of challenges comes into play when different notions of solution are discussed. They are highlighted, and to a large extent addressed, by the Yamada-Watanabe Theorem. For the sake of clarity, we discuss the notions of solution, different notions of uniqueness and the basic statement of the Yamada-Watanbe Theorem for SDEs in Section 2.3. The general importance of Yamada-Watanabe results can be highlighted by an example of a typical application. The diffusion approximation of population models is quite often obtained by construction of a sequence of approximations which corresponds to a sequence of martingale problems. Then, after tightness of the sequence is established, one concludes that the limit point of this sequence describes the diffusion approximation. This strategy is used in, for example, Etheridge et al. (2014) and in Chapter 3. However, this procedure leads to two important questions.

The first one is the question of uniqueness of solutions to the limiting martingale problem. In many situations this question is addressed by the means of duality (e.g. Etheridge et al. (2014)), or approximate duality (see Mytnik (1996b)). However, sometimes the duality relation is not available. This is precisely the situation that we encounter in the case of the Spatial Lambda-Fleming-Viot model with selection in a fluctuating environment in Chapter 3. In this situation, one may infer the uniqueness of solutions to the martingale problem from the weak uniqueness of the corresponding SPDE (provided that these two are equivalent), which, in turn, can be deduced from a pathwise uniqueness result for the SPDE, as the existence is guaranteed by the approximation procedure.

The second question has a more universal nature. Even if the uniqueness of solutions to the martingale problem can be established, this type of uniqueness guarantees only weak uniqueness of solutions of the corresponding S(P)DE. This guarantees only that there exists a probability space on which the processes are defined and they satisfy the limiting equation. If pathwise uniqueness can be shown, the Yamada-Watanabe Theorem allows us to deduce existence and uniqueness in the strong sense - this means that the question of uniqueness can be answered affirmatively on every probability space that can support the processes.

The rest of the chapter is organised as follows. In Section 2.1 we recall some basic facts on trace class operators and Hilbert-Schmidt operators and we define Gaussian

measures and Gaussian processes on Hilbert spaces. In Section 2.2 we define two different stochastic integrals, which help to give a meaning to SHE (2.1) and SHE (2.2). We begin by defining the integrators in Subsection 2.2.1, where we discuss the Q-Wiener measure, cylindrical Brownian motion and martingale measures and some connections between them. In Subsection 2.2.2 we sketch the constructions of integrals with respect to the defined integrators. Subsection 2.2.3 contains information on some more recent developments in extending the integrals we have defined. In Section 2.3 we discuss the classical Yamada-Watanabe Theorem. In Section 2.4 we describe a more general framework in which a Yamada-Watanabe-Theorem can be proved and use it to prove a Yamada-Watanabe Theorem for solutions of SHE. In Section 2.5 we state useful results on SHE, including an existence result and a pathwise uniqueness result. This section also contains a proof that the solution of a martingale problem and a weak solution of the related SPDE are equivalent. Section 2.7 contains some facts on Poisson random measures and conditionally Poisson systems.

2.1 Some facts from functional analysis

In this subsection we recall some facts from functional analysis which form the backbone of our constructions. We begin with a short description of trace class operators. This class of operators allows us to define an integral with respect to infinite dimensional noise. There are many natural norms for the trace class operators. Two of them are particularly useful - the trace norm, $\|\cdot\|_{Tr}$, which is an analogue of the usual L^1 norm, and the Hilbert-Schmidt norm, $\|\cdot\|_{HS}$, which is an analogue of the L^2 norm. Both play an important role in infinite dimensional stochastic analysis.

We briefly introduce the trace class in Subsection 2.1.1. Then, in Subsection 2.1.2, we discuss the Hilbert-Schmidt class of operators. We conclude this section with a brief description of Gaussian measures on Hilbert spaces in Subsection 2.1.3. The results presented are standard and we point towards the relevant references along the way.

2.1.1 Trace class operators

An excellent introductory treatment of this topic can be found in Lax (2002), Chapter 30.

Definition 2.1 (Trace class operator)**.** *Let $T : H \to H$ be a compact operator mapping the Hilbert space H onto itself. Let λ_i^T denote the set of its non-zero eigenvalues. We say that the operator T is in the trace class if its trace norm, $\|\cdot\|_{Tr}$, is finite, that is*

$$\|T\|_{Tr} = \sum_i \lambda_i^T < \infty.$$

The trace class is closed under adjoints and the trace norm satisfies the triangle inequality. We provide an alternative perspective on trace class operators, using an analogue of the finite dimensional notion of trace. Every bounded operator T on a Hilbert space H can be represented as an infinite dimensional matrix with respect to any orthonormal basis f_n. The mnth element of this matrix is (Tf_nf_m). The trace of this matrix is

$$Tr(T) = \sum_i \langle Tf_i, f_i \rangle_H, \tag{2.6}$$

as long as this series converges. The importance of the trace is best explained by the next two theorems.

Theorem 2.2 (Trace invariance)**.** *For every element of the trace class, the series* (2.6) *converges. What is more, the limit is invariant under change of orthonormal basis.*

The next Theorem, due to Lidskiĭ (1959), links the trace norm and the trace.

Theorem 2.3 (Lidskiĭ's trace formula)**.** *For every trace class operator T, the trace is given by the sum of its eigenvalues, that is*

$$Tr(T) = \sum_i \lambda_i^T.$$

In other words, for an operator from the trace class, the trace and the trace norm are equal.

2.1.2 Hilbert-Schmidt operators

In the construction of integrals in infinite dimensional spaces, a Hilbert space of a class of operators bearing a strong similarity to the trace class operators is particularly useful. This subsection contains a definition of the nuclear operators and Hilbert-Schmidt operators. For the proofs, and most of the background material from an analytic perspective, we refer to Young (1988), Chapter 8. In this subsection, H_1 and H_2 are always Hilbert spaces, H_1^*, H_2^* denote their respective duals and $L(H_1, H_2)$ denotes the space of the linear maps between H_1 and H_2.

We begin by defining an analogue of trace class operators for maps defined between two Hilbert spaces.

Definition 2.4 (Nuclear operators)**.** *Let $T \in L(H_1, H_2)$. We say that T is a nuclear operator if for each $x \in H_1$, T can be represented as*

$$Tx = \sum_i a_i \phi_i(x),$$

where $a_j \in H_2$ and $\phi \in H_1^*$ satisfy

$$\sum \|a_i\|_{H_2} \|\phi_i\|_{H_1^*} < \infty.$$

We observe that if $H_1 = H_2$, the nuclear operators are just trace class operators.

Definition 2.5 (Hilbert-Schmidt operator)**.** *Let H_1, H_2 be a pair of Hilbert spaces. We say that a bounded linear operator $T : H_1 \to H_2$ is Hilbert-Schmidt if for any orthonormal basis of the space H_1, e_k,*

$$\sum_i \|Te_i\|_{H_2}^2 < \infty. \tag{2.7}$$

It turns out that the expression (2.7) (which is independent of the choice of the basis) defines a square of a norm on the space of nuclear operators.

Definition 2.6 (Hilbert-Schmidt space)**.** *Let $L_{HS}(H_1, H_2)$ denote the space of all operators for which expression* (2.7) *is finite. We call this space the space of Hilbert-Schmidt operators. For every operator T, we define the Hilbert-Schmidt norm, $\|\cdot\|_{HS}$*

as

$$\|T\|_{HS} = \left(\sum_i \|Te_i\|_{H_2}^2 \right)^{\frac{1}{2}} < \infty. \tag{2.8}$$

To see that $\|\cdot\|_{HS}$ is independent of the choice of basis we note that

$$\|T\|_{HS}^2 = \sum_i \|Te_i\|_{H_2}^2 = \sum_i \sum_j \langle Te_i, f_j \rangle = \sum_i \|T^* f_i\|,$$

where f_j is the basis of H_2^*. This computation shows also that the adjoint of a Hilbert-Schmidt operator is a Hilbert-Schmidt operator. The space of Hilbert-Schmidt operators between H_1 and H_2, which we denote by $L_2(H_1, H_2)$ endowed with the norm $\|\cdot\|_{HS}$ is a Hilbert space with a scalar product given by

$$\langle T, S \rangle_{HS} = \sum_i \langle Te_i, Se_i \rangle_{H_2}$$

We draw further connections between Hilbert-Schmidt operators and trace class operators.

Theorem 2.7. *Let T and K be the trace class operators from Hilbert space H taking values in H.*

1. *If $T \in L_2(H, H)$, the Hilbert Schmidt norm can be expressed as*

$$\|T\|_{HS}^2 = \sum_i \|Te_i\|_H^2 = \sum_i (\lambda_i^T)^2,$$

   where λ_i^T is the collection of eigenvalues of the operator T,

2. *Every trace class operator is Hilbert-Schmidt,*

3. *Every trace class operator is a product of two Hilbert-Schmidt operators.*

2.1.3 Gaussian measures and processes on Hilbert spaces

This subsection is based on Da Prato and Zabczyk (2008), Chapter 2. In the finite dimensional setup, the basic building block for the theory of stochastic integration is

Brownian motion, which is a Gaussian process. Here we would like to define a Gaussian process on a Hilbert space.

Definition 2.8. *Let H be a Hilbert space. A probability measure μ, defined on all Borel sets $\mathcal{B}(H)$ of the space H, is a Gaussian measure if every linear functional $h \in H$, considered as a random variable on $(H, \mathcal{B}(E), \mu)$, is a Gaussian measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. In other words, μ is Gaussian if, for each $h \in H$, there exists a pair of real numbers $a \in \mathbb{R}$ and $b \in \mathbb{R}_+ \cup \{0\}$ such that for every set $A\mathcal{B}(\mathbb{R})$*

$$\mu\{x \in H : \langle h, x\rangle \in A\} = \mathcal{N}(a, b)(A),$$

where $\mathcal{N}(a, b)$ denotes a normal distribution with mean a and variance b.

We specify the analogues of mean and variance for the Hilbert space-valued Gaussian measures. The definition of a Gaussian measure on the Hilbert space H implies that for every Gaussian measure μ, there exists $m \in H$ such that

$$\int_H \langle h, x\rangle_H \mu(\mathrm{d}x) = \langle h, m\rangle.$$

We call m the mean value of μ. Analogously, there exists a linear operator Q such that for any pair $h_1, h_2 \in H$

$$\int_H \langle h_1, x - m\rangle_H \langle h_2, x - m\rangle_H \mu(\mathrm{d}x) = \langle Qh_1, h_2\rangle_H.$$

We call Q the covariance operator. It is clear from the definition that it is symmetric and nonnegative. We know that in finite dimensions, the Gaussian measure is uniquely determined by its mean and covariance. An analogous result holds for Gaussian measures on Hilbert spaces.

Theorem 2.9. *Let μ be a Gaussian measure on a Hilbert space H. Then μ is uniquely determined by its mean m and covariance operator Q.*

The following two statements highlight the importance of trace class operators for the theory of Gaussian processes on Hilbert spaces. The first one states that the trace class operators are the only possible covariance operators.

Theorem 2.10 (Da Prato and Zabczyk (2008), Proposition 2.16)**.** *Let μ be a Gaussian measure on a Hilbert space H with mean $m = 0$ and covariance operator Q. Then Q belongs to the trace class on H.*

The second theorem states that every trace class operator corresponds to a Gaussian measure on a Hilbert space.

Theorem 2.11 (Da Prato and Zabczyk (2008), Proposition 2.18)**.** *Let Q be a positive, symmetric trace class operator on H. Let $m \in H$. There exists a Gaussian measure with covariance operator Q and mean m.*

We close this section with a definition of a Gaussian process on a Hilbert space. It is completely analogous to its finite dimensional counterpart.

Definition 2.12 (Gaussian process)**.** *We say that a stochastic process X, taking values in a Hilbert space H, is Gaussian if for any finite collection $(t_1, \ldots, t_n)$ the random variable $(X(t_1), \ldots, X(t_n))$ is Gaussian; that is, if every finite dimensional distribution of the process is Gaussian.*

2.2 Stochastic integral

The main aim of this section is to provide a very brief sketch of two constructions of stochastic integrals which allow us to give a meaning to solutions of SPDE. We begin with a description of the stochastic processes which we use as integrators. This is the main content of Subsection 2.2.1. We provide two different approaches to giving a meaning to the integral with respect to an infinite dimensional stochastic process and point to connections between them. This is the content of Subsection 2.2.2. Finally, in Subsection 2.2.3, we discuss some extensions of the presented theory.

2.2.1 Integrators

In this subsection we provide a short description of the class of processes with respect to which we would like to integrate. First, we describe the Q-Wiener measure and the cylindrical Brownian motion, which correspond to the coloured and white noise respectively. These two objects are natural extensions of the usual Brownian motion

in the context of Hilbert (and Banach) spaces, and correspond to the interpretation of solutions to (2.2). The other class of integrators used in this thesis is the class of worthy martingale measures. These objects are natural extensions of finite dimensional martingales. Integrals with respect to martingale measures correspond to the random field interpretation of a solution to equation (2.1).

2.2.1.1 Q-Wiener measure

We begin with a definition of Q-Wiener measure on a Hilbert space H. It is a direct extension of the definition of the usual Brownian motion. The natural problems arising from the infinite dimensional nature of the object are addressed by referring to properties of trace class operators.

Definition 2.13 (Q-Wiener process)**.** *Let H be a Hilbert space and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete (that is, containing all $\mathbb{P}$-null sets) probability space. Let Q be a continuous, linear, nonnegative, symmetric trace class operator. We say that an H-valued stochastic process W is a Q-Wiener process if*

1. *$W(0) = 0$,*
2. *the trajectories of W are almost surely continuous,*
3. *the increments of W are independent, that is, for $0 \leq t_1 < t_2 < \cdots < t_n < T$ the random variables*

$$W(t_1), W(t_2) - W(t_1), \ldots, W(t_n) - W(t_{n-1})$$

   are independent,

4. *the increments of W have Gaussian distribution with mean 0 and covariance Q, that is*

$$\mathbb{P}[W(t) - W(s)]^{-1} = \mathcal{N}(0, (t-s)Q).$$

Theorem 2.14 (Real-valued representation of the Q-Wiener process, Da Prato and Zabczyk (2008), Proposition 4.3 and Theorem 4.5)**.** *Let H be a Hilbert space. Let Q*

be a continuous, linear, nonnegative, symmetric operator with finite trace. Let e_k be a collection of eigenvectors of Q together with their eigenvalues λ_k. Then an H-valued stochastic process W is a Q-Wiener process if and only if

$$W(t) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} W_t^i e_i, \tag{2.9}$$

where $\{W^i\}_{i=1}^{\infty}$ is a collection of independent standard Brownian motions, and the series (2.9) *converges in probability.*

Since by Definition 2.13 a Q-Wiener process is continuous, so is the process from Theorem 2.14.

2.2.1.2 Cylindrical Brownian motion

The processes of Definition 2.13 provide us with a broad class of processes which form an infinite dimensional analogue of the standard Brownian motion. However, not all of the natural extensions of the Brownian motion of interest are of this form. If we take $Q = Id$, where Id is the identity operator, the resulting object leads to a noise that we have intuitively described as white noise. The identity operator is, of course, not in the trace class. It is therefore desirable to extend the definition of the infinite-dimensional Wiener process to the case where the operator Q does not belong to the trace class.

There is more than one approach which makes this extension possible. The first one, described in detail in Da Prato and Zabczyk (2008) Section 4.1.2, is to consider a generalised Wiener process, defined for every $a \in H$ as

$$W_a(t) = \langle a, W(t) \rangle_H,$$

where $\langle \cdot, \cdot \rangle_H$ denotes the scalar product in H. We note that the definition makes sense if the map $a \mapsto W_a$ is linear and valued in $\mathbb{R}$. This is the case as long as we can find a continuous, positive, self-adjoint operator Q such that for any $t, s \geq 0$ and $a, b \in H$

$$\mathbb{E}[W_a(t) W_b(s)] = (t \wedge s) \langle Qa, b \rangle_H. \tag{2.10}$$

Definition 2.15. *We say that W is a generalised Wiener process with covariance Q whenever W is a centred Gaussian process with covariance structure described by* (2.10)*. If $Q = Id$, the identity operator, then we call this Wiener process a cylindrical Wiener process.*

Alternatively, we may follow Liu and Röckner (2015). Define a Hilbert-Schmidt embedding $J : H \to H$, mapping the Hilbert space H into itself, by

$$J(a) = \sum_i \alpha_i \langle a, e_i \rangle_H e_k,$$

where e_i is an orthonormal basis of H and $\{\alpha_i\}_{i=1}^{\infty}$ is a square summable sequence of real numbers. Observe that the map J is injective. Then, if W^i is a sequence of independent standard one dimensional Brownian motions,

$$W(t) = \sum_i W^i(t) J(e_i)$$

defines a Q_1-Wiener measure on H, where $Q_1 = JJ^*$. In this way we overcome the main difficulty of dealing with a covariance operator which is not in the trace class - we express it on a different Hilbert space (which we can choose to be a copy of the initial one), but now due to the action of the Hilbert-Schmidt operator our object is in the trace class. This approach simplifies the definition of the integral with respect to the cylindrical Brownian motion which follows later. Notice that the same operation could have been performed for the usual Q-Wiener process, however there would be no gain in doing so. For the details, we refer to Liu and Röckner (2015), Section 2.5.1, especially Proposition 2.5.2.

2.2.1.3 Worthy martingale measures

In this subsection we make the first step towards defining the integral which allows us to give meaning to solutions of equation (2.1) by defining the integrands. This approach to constructing an infinite dimensional stochastic integral is closest in spirit to the construction of the stochastic integral with respect to finite dimensional continuous martingales. We are interested in objects which have both spatial and temporal

components. This requires a construction of random measures. Ideally, we would like the random measures to be countably additive, but for most examples of interest this cannot be achieved. Fortunately, it is enough to demand the countable additivity only in an L^2 sense, which could be expected from the fact that in the finite dimensional setting one defines the integral for square integrable martingales. This subsection is based mostly on Walsh (1986).

Definition 2.16 (σ-finite L^2 measure)**.** *Consider a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ and a Polish space E, that is, a separable, completely metrizable, topological space. Let $\mathcal{B}(E)$ denote the Borel algebra on E. Consider a random set function $U : \mathcal{B}(E) \times \Omega \to \mathbb{R}$ which satisfies the following conditions:*

1. *U is finitely additive, that is, for any two disjoint sets $A, B \in \mathcal{B}(E)$*

$$U(A \cup B) = U(A) + U(B),$$

2. *U is bounded in $L^2(E \times \Omega)$, that is, for each $A \in \mathcal{B}(E)$*

$$\mathbb{E}\Big[U^2(A)\Big] < \infty$$

3. *for any collection of non-increasing Borel sets with empty intersection, that is $A_{j+1} \subset A_j$, and $\bigcap A_j = \emptyset$*

$$\lim_{j \to \infty} \mathbb{E}\Big[U^2(A_j)\Big] = 0.$$

If in addition U is a σ-finite on E, we say that U is σ-finite L^2 measure.

This definition allows us to characterize the (worthy) martingale measures, which will form our class of integrators.

Definition 2.17 (Martingale measure)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $\mathcal{F}_t$ a complete right continuous filtration. Let E be a Polish space. Consider a stochastic process $M_t : E \times (0, \infty) \to \mathbb{R}$. We say that M_t is a martingale measure if*

1. *for each $A \in \mathcal{B}(E)$, $M_0(A) = 0$,*

2. for each $A \in \mathcal{B}(E), M_t(A)$ is a martingale,

3. for every $t \geq 0$, M_t is a σ-finite L^2 measure.

In analogy to the theory of finite dimensional martingales, we define the bracket process for a martingale measure.

Definition 2.18 (Covariance functional)**.** *Let M be a martingale measure. Consider Borel sets $A, B \in \mathcal{B}(\mathbb{R}^d)$. We say that a biadditive, random, symmetric set function Q_t^M is the covariance functional of M if*

$$Q_t^M(A, B) = \langle M(A), M(B) \rangle_t.$$

With a slight abuse of notation we observe that

$$Q^M(A \times B \times (s, t]) = Q_t^M(A, B) - Q_s^M(A, B)$$

defines an associated covariance measure.

Unfortunately, not all martingale measures give rise to well defined integrators. A neat counterexample has been given by Bakry, see Section 3 of Walsh (1986), pp. 306-308 and Bakry (1982). We therefore restrict the class of integrators to worthy martingale measures.

Definition 2.19 (Worthy martingale measure)**.** *Let M_t be a martingale measure. If there exists a positive Radon measure (locally finite, inner regular measure) K on $E \times E \times [0, \infty)$ such that for every $A, B \in \mathcal{B}(E)$ and each $t, s \in \mathbb{R}_+$*

$$|\langle M(A), M(B) \rangle_{t+s} - \langle M(A), M(B) \rangle_t| \leq K[A \times B \times (t, t+s]] < \infty,$$

we say that the martingale measure M_t is worthy. We call the measure K a dominating measure.

Before we provide some examples, let us describe a condition which allows to easily verify existence of the dominating measure. We need one more definition.

Definition 2.20 (Orthogonality)**.** *We say that a random measure M is orthogonal if for each pair of disjoint sets $A, B \in E$ and each $t > 0$*

$$\langle M(A), M(B)\rangle_t = 0.$$

Lemma 2.21. *If a martingale measure M is orthogonal, it is a worthy martingale measure.*

Proof Let $\pi : \mathcal{B}(E) \times [0,\infty) \to \mathbb{R}$ be defined as $\pi(A,[0,t)) = \langle M(A)\rangle_t$. Let $f : E \times [0,\infty \to E \times E \times [0,\infty)$ be defined as $f(x,t) = (x,x,t)$. If for $\Gamma \subset E \times E \times [0,\infty)$ we set

$$K(\Gamma) = \pi(f^{-1}(\Gamma)),$$

K is a dominating measure for M. Then, since M is a martingale measure, it is worthy. □

We are now ready to provide some examples of worthy martingale measures. We define a worthy martingale measure which corresponds to the object we have described as white noise in the beginning of this chapter. Let Λ be standard Lebesgue measure. White noise is a real valued random set function $\mathcal{W}$ on finite measure Borel sets which satisfies

1. $\mathcal{W}(A)$ is a Gaussian random variable with mean 0 and variance $\Lambda(A)$,
2. if A and B are disjoint, then $\mathcal{W}(A)$ and $\mathcal{W}(B)$ are independent.
3. for any Borel sets A, B the covariance function is given by

$$\mathbb{E}[\mathcal{W}(A)\mathcal{W}(B)] = \Lambda(A \cap B).$$

It is easy to construct a worthy martingale measure from $\mathcal{W}$. Observe that for $A \in \mathcal{B}(\mathbb{R}^d)$ of finite measure and $t > 0$,

$$M(A,t) = \mathcal{W}(A \times [0,t])$$

defines an orthogonal martingale measure. Therefore, by Lemma 2.20, it defines a worthy martingale measure.

The second example corresponds to an object which we have intuitively described as coloured noise. Here we briefly sketch the construction of a coloured noise with correlation function $q(x,y) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and show why it gives a rise to a worthy martingale measure. We assume that the function q is symmetric, continuous and locally integrable.

Consider the Schwartz space $S(\mathbb{R}^{d+1})$ of rapidly decaying functions on $\mathbb{R}^{d+1}$, that is the space

$$S(\mathbb{R}^{d+1}) = \{f \in C^\infty(\mathbb{R}^n) : \|f\|_{\alpha,\beta} < \infty \quad \forall \alpha, \beta \in \mathbb{N}^n\},$$

where α, β are multi-indices and the seminorm $\|\cdot\|_{\alpha,\beta}$ is defined as

$$\|f\|_{\alpha,\beta} = \sup_{x \in \mathbb{R}^{d+1}} \left| x^\alpha D^\beta f(x) \right|.$$

It is well known that this space is dense in L^2 and its dual is the space of tempered distributions. Let $\langle \cdot, \cdot \rangle_{L^2(\mathbb{R}^d)}$ denote the scalar product on $L^2(\mathbb{R}^d)$.

For any pair of functions $\phi(s,x), \psi(s,x) \in S(\mathbb{R}^{d+1})$, we may now define a bilinear form J_q by

$$J_q(\phi,\psi) = \int_{\mathbb{R}} \langle Q_g \phi(s,\cdot), \psi(s,\cdot) \rangle_{L^2(\mathbb{R}^d)} \mathrm{d}s = \int_{\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}^d} \phi(s,x) q(x,y) \psi(s,y) \mathrm{d}x \mathrm{d}y \mathrm{d}s$$

Our $q(x,y)$ is chosen in a way that ensures that J_q is symmetric, continuous and positive definite.

Since J_q is a Hilbert-Schmidt operator on $C_c^\infty(\mathbb{R}^{d+1})$, by Theorem 2.11 it uniquely determines a mean zero Gaussian measure on the space of regular distributions. We may further extend this Gaussian measure to a martingale measure on $L^2(\mathbb{R}^{d+1})$ by a density argument (as every function of the form $\mathbf{1}_A(x)$, where $\mathbf{1}$ is the set indicator function, can be approximated by $C_c^\infty(\mathbb{R}^d)$ functions). We define the martingale measure M_q

with covariance q by

$$M_q(A,t) = W(A, \times[0,t]),$$

where for any pair of Borel sets A, B, W satisfies

1. $W(A)$ is a Gaussian random variable with mean 0 and variance $\int_A \int_A q(x,y)\mathrm{d}x\mathrm{d}y$,
2. if A and B are disjoint, then $W(A)$ and $\mathcal{W}(B)$ are independent.
3. the covariance function is given by

$$\mathbb{E}[W_s(A)W_t(B)] = (s \wedge t) \int_A \int_B q(x,y)\mathrm{d}x\mathrm{d}y. \tag{2.11}$$

For the details of the construction, we refer to Walsh (1986), Section 2.

To see that the constructed martingale measure is worthy, it is enough to observe that the covariance measure

$$\mathbb{E}[M_s(A)M_t(B)] = (s \wedge t) \int_A \int_B q(x,y)\mathrm{d}x\mathrm{d}y$$

gives rise to a well defined dominating measure.

To demonstrate that the notion of a worthy martingale measure is not limited to Gaussians, we sketch the construction of a worthy martingale measure corresponding to the Poisson random measure. For simplicity, we consider a Poisson random measure defined on $\mathbb{R}$.

Definition 2.22 (Poisson random measure)**.** *We say that N is a Poisson random measure on $\mathbb{R} \times [0,\infty)$ with mean measure $\nu \times \Lambda$ (where Λ denotes a standard Lebesgue measure) if for every pair of Borel sets $A, B \in \mathcal{B}(\mathbb{R} \times [0,\infty))$,*

1. *$N(A)$ is a Poisson random variable with mean $(\nu \times \Lambda)(A)$,*
2. *if $A \cap B = \emptyset$, $N(A)$ and $N(B)$ are independent.*

Definition 2.22 can be easily extended to any Polish space E instead of $\mathbb{R}$.

Consider a Poisson random measure with mean measure $\nu \times \Lambda$, and a function $h \in L^2(\mathbb{R}, \nu)$, that is a function on $\mathbb{R}$ that is square integrable with respect to the measure ν. Then, for every Borel set $A \in \mathcal{B}(\mathbb{R} \times [0, \infty)$, we define M as

$$M(A, t) = \int_A h(x)[N(\mathrm{d}x \times [0, t)) - t\nu(\mathrm{d}x)].$$

Then M is an orthogonal martingale measure, and therefore a worthy martingale measure. The dominating measure K can be chosen as follows:

$$K(A \times B \times [0, t)) = \langle M(A), M(B)\rangle_t = t \int_{A \cap B} h^2(x)\mathrm{d}x.$$

2.2.1.4 Gaussian coloured noise as a Q-Wiener measure

We would like to show a link between the Q-Wiener measure and coloured noise with correlation function $q(x, y)$. If the correlation kernel $q(x, y)$ is a square integrable function (i.e. $q(x, y) \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$), the coloured noise is a Q-Wiener measure with the covariance functional Q determined by an orthogonal decomposition of $q(x, y)$ in $L^2(\mathbb{R}^d \times \mathbb{R}^d)$. Let us explain in a little bit more detail.

By Theorem 2.14, the covariance functional of the Q-Wiener measure can be represented as

$$\mathbb{E}[\langle W_t, \phi\rangle, \langle W_t, \psi\rangle] = \sum_{i,j}^{\infty} \mathbb{E}[W_t^i W_t^j]\sqrt{\lambda_i}\sqrt{\lambda_j}\langle \phi, e_i\rangle\langle \psi, e_j\rangle.$$

Since W_t^i and W_t^j are independent, this can be simplified to

$$t\sum_i^{\infty} \lambda_i \langle \phi, e_i\rangle\langle \psi, e_i\rangle = \sum_i^{\infty} t \int_{\mathbb{R}\times\mathbb{R}^d} \sqrt{\lambda_i}\phi(x)e_i(x)\sqrt{\lambda_i}\psi(y)e_i(y)\mathrm{d}x\mathrm{d}y.$$

We observe that this expression is the same as (2.11), provided that

$$q(x, y) = \sum_i^{\infty} \sqrt{\lambda_i}e_i(x)\sqrt{\lambda_i}e_i(y),$$

and the series is square summable.

The set of conditions that guarantee existence of such a representation of the correlation kernel is given in the following theorem, sometimes named after Mercer.

Theorem 2.23 (Mercer's Theorem, Young (1988), Theorem 8.8)**.** *Let $q(x,y) \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$ be symmetric. Then there exists a collection of orthogonal functions $f_i \in L^2(\mathbb{R}^d)$ such that*

$$q(x,y) = \sum_i f_i(x) f_i(y)$$

and this series converges in $L^2(\mathbb{R}^d \times \mathbb{R}^d)$.

In order to recover the Q-Wiener measure (or construct the correlation kernel) it is enough to take $f_i = \sqrt{\lambda_i} e_i$.

This discussion allows us to observe that the following representation is valid:

$$W(t,\phi) = \int_{\mathbb{R}^d} \phi(x) W(\mathrm{d}t, \mathrm{d}x) = \sum_{i=1}^{\infty} W_t^i \int_{\mathbb{R}^d} f_i(x)\phi(x)\mathrm{d}x. \tag{2.12}$$

We underline that application of Theorem 2.23 gives rise to a wide class of examples of correlation kernels.

2.2.2 Stochastic integral

In this subsection, we sketch the construction of an integral with respect to Q-Wiener measures, cylindrical Brownian motion and worthy martingale measures.

2.2.2.1 Integral with respect to Q-cylindrical Wiener measure

The exposition in this subsection follows Da Prato and Zabczyk (2008), Chapter 4, and Prévôt and Röckner (2007).

In this subsection we work with a Hilbert space H. We also assume that the Q-Wiener process under consideration is defined as in Definition 2.13. In particular, this means that the process is measurable with respect to a complete filtration $\mathcal{F}_t$.

The construction of the stochastic integral follows the usual procedure. First we define the integral for the set of elementary processes for which it is clear that the integral makes sense. We later show that the closure of the set of elementary processes

is large enough to give rise to a reasonably rich space of integrands. Unsurprisingly the class of elementary processes is defined in a very similar way to the finite dimensional setting.

Definition 2.24 (Elementary process)**.** *Fix $T > 0$. We say that the process $\Phi(t)$, taking values in the space of bounded linear operators from H to H, is an elementary process if there exists a sequence $0 = t_0 < t_1 \cdots < t_{n-1} < t_n = T$ and a sequence of L^1-valued random variables $\{\Phi_i\}_{i=0}^n$ such that Φ_i is $\mathcal{F}_{t_i}$ measurable and*

$$\Phi(t) = \Phi_i \quad \text{for } t \in [t_i, t_{i+1}).$$

It is easy to define the integral with respect to the elementary processes, which we denote by $\Phi \cdot W$.

Definition 2.25 (Integral with respect to elementary process)**.** *Let Φ be an elementary process. The stochastic integral of Φ with respect to a Q-Wiener measure W is defined as*

$$\Phi \cdot W = \int_0^T \Phi(s)\mathrm{d}W(s) = \sum_{i=1}^{k-1} \Phi_i(W_{t_{i+1}\wedge t} - W_{t_i \wedge t}).$$

This definition leads to an object which is an analogue of the integral of an elementary process with respect to a Brownian motion. This is illustrated by the following proposition, where we make use of the theory of Hilbert-Schmidt operators, as described in Subsection 2.1.2.

Theorem 2.26 (Da Prato and Zabczyk (2008), Proposition 4.20)**.** *Let Φ be an elementary process. Then $\Phi \cdot W$ is a continous square integrable martingale, and*

$$\mathbb{E}\Big[(\Phi \cdot W)^2\Big] = \mathbb{E}\Big[\int_0^t \|\Phi\|_{HS}^2\Big]. \tag{2.13}$$

It turns out that the right choice of the class of integrands is the class of predictable processes with values in the space of Hilbert-Schmidt operators. For the sake of completeness, we recall the definition of predictability.

Definition 2.27 (Predictable processes)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a complete probability space and let $\mathcal{F}_t$ be a complete filtration. Let $\mathcal{P}_T$ denote the σ-field generated by sets of the form*

$$(s,t] \times F, \quad 0 < s < t < T, F \in \mathcal{F}_s.$$

We say that $\mathcal{P}_T$ is the predictable σ-field and its elements are predictable sets. The process is called predictable if it is measurable with respect to $\mathcal{P}_T$.

Remark 2.28. *We would like to make a remark about the relationship between adapted and predictable processes.*

1. *If a process is predictable, it is adapted.*
2. *If a process is adapted and continuous, there exists a modification of the process which is predictable.*

The following theorem allows us to extend the definition of the stochastic integral.

Theorem 2.29 (Da Prato and Zabczyk (2008), Proposition 4.22.)**.** *The following are true:*

1. *The elementary processes taking values in the Hilbert-Schmidt class are predictable.*
2. *If Φ is a predictable process such that*

$$\mathbb{E}\left[\int_0^t \|\Phi\|_{HS}^2\right] < \infty,$$

then there exists a sequence of elementary functions Φ_n such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^t \|\Phi - \Phi_n\|_{HS}^2\right] \to 0.$$

Remark 2.30. *It is possible to further extend the definition of the integral to all processes which satisfy*

$$\mathbb{P}\left[\int_0^t \|\Phi\|_{HS}^2 \mathrm{d}s < \infty\right] = 1.$$

For the details, we refer to Da Prato and Zabczyk (2008), Lemma 4.24

The integral defined in this subsection has all the properties we would expect - the result of integration with respect to a martingale gives us a martingale, a stochastic equivalent of Fubini Theorem is true and Itô's formula is available.

So far we have presented a theory of integration which applies to Q-Wiener processes. We now describe the extension to Gaussian processes with covariance operator Q outside of the trace class, as described in Subsection 2.2.1.2.

Recall that we have defined a Hilbert-Schmidt embedding $J : H \to H$, leading to a Q_1-Wiener process, with $Q_1 = JJ^*$. Since the new process is given by

$$W'(t) = \sum_i W^i(t) J e_i,$$

and its covariance is in the trace class, our definition of integral applies. To recover the integral with respect to the original process it is enough to evaluate

$$\int_0^t \Phi(s) \mathrm{d}W(s) = \int_0^t \Phi(s) \circ J^{-1} \mathrm{d}W'(s),$$

where $\circ$ denotes composition of functions. For details, we refer to Liu and Röckner (2015), Section 2.5.2.

2.2.2.2 Integral with respect to a worthy martingale measure

The exposition in this subsection follows Walsh (1986) and Dalang (1999).

In a similar way to the finite-dimensional Itô integral and the integral with respect to a Q-Wiener measure, the construction of an integral with respect to a worthy martingale measure starts with a choice of a large enough class of elementary processes, for which the definition of the integral is simple. Then one extends the construction by approximating the larger class of processes with elementary processes, using an appropriate norm. We briefly sketch the elements of this construction.

Consider a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a complete filtration $\mathcal{F}_t$. We say that a function $f : [0, \infty) \times \mathbb{R}^d \times \Omega \to \mathbb{R}$ is an elementary function if for

$0 \leq s \leq t$, a Borel set A and a bounded, $\mathcal{F}_t$ measurable process $X(t,\omega)$, it can be written as

$$f(s,x,\omega) = 1_{(a,b]}(s)1_A(x)X(\omega),$$

for some $a, b \in \mathbb{R}_+$ We define the integral of such a function with respect to a worthy martingale measure in the expected way, that is, we measure the increase of the value of the martingale over a fixed Borel set over $[s,t)$.

Definition 2.31 (Integral with respect to an elementary function)**.** *Let f be an elementary function. Let M be a worthy martingale measure. The integral of f with respect to M over a Borel set B, which we denote by $(f \cdot M)_t(B)$, is defined as*

$$(f \cdot M)_t(B) = \int_0^t \int_B f(s,x,\cdot) M(\mathrm{d}s\mathrm{d}x) = \Big(M_{a\wedge t}(A \cap B) - M_{b\wedge t}(A \cap B) \Big) X.$$

Definition 2.31 can be extended in an obvious way to the set of all linear combinations of elementary functions. We denote this set by $\mathcal{E}$.

Now we move to the second part of the procedure. We would like to find a norm in which the closure of the set $\mathcal{E}$ is large enough to define a reasonably large class of integrands. Such a norm, which we denote by $\|\cdot\|_+$, can be defined as

$$\|f\|_+ = \left(\mathbb{E}\left[\int_0^t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(s,x,\cdot)||f(s,y,\cdot)| K_M(\mathrm{d}x\mathrm{d}y\mathrm{d}s) \right] \right)^{\frac{1}{2}},$$

where K_M denotes the dominating measure for a worthy martingale M. It is easy to check that $\|\cdot\|_+$ does indeed define a norm. Let $\mathcal{P}_+^M$ denote the space of all predictable functions f for which the norm $\|\cdot\|_+$ is finite. This space is complete. This a consequence of two observations. First, it is easy to see that boundedness in $\|\cdot\|_+$ guarantees that the map $\overline{f}$ defined by

$$\overline{f}(s,x,y,\omega) = f(s,x,\omega) f(s,y,\omega) \tag{2.14}$$

is integrable with respect to the measure $K_M(\mathrm{d}x\mathrm{d}y\mathrm{d}s)\mathbb{P}(\Omega)$. It is therefore bounded in L^1, which is complete. Then it is enough to observe that the limits of functions of the

form (2.14) are also of the same form. Therefore the completeness of L^1 translates into completeness of $\mathcal{P}_+^M$.

We now state the theorem which allows us to extend the integral to all elements of $\mathcal{P}_+^M$.

Theorem 2.32 (Walsh (1986), Proposition 2.3)**.** *The set of elementary functions $\mathcal{E}$ is dense in $\mathcal{P}_+^M$ with respect to the norm $\|\cdot\|_+$.*

It is now straightforward to extend the definition of the integral to the class of predictable processes. By Theorem 2.32 for each $f \in \mathcal{P}_+^M$ we can find a sequence of elementary functions $f_n \in \mathcal{E}$ such that $\|f - f_n\|_+ \to 0$. Since for every Borel set A and every $f \in \mathcal{E}$

$$\mathbb{E}\Big[f \cdot M_t^2(A)\Big] \leq \|f\|_+^2,$$

the sequence $f_n \cdot M_t(A)$ is Cauchy in $L^2(\Omega)$ and defines a unique limit. We define the integral $f \cdot M_t(A)$ to be equal to this limit.

The integral with respect to a worthy martingale measure as defined in this subsection has many desirable properties. In complete analogy to the standard Itô integral, $f \cdot M$ defines a worthy martingale measure (and therefore a new process with respect to which we can integrate), with the expected covariance structure, see Walsh (1986), Theorem 2.5. One can also prove an analogue of the classical Fubini Theorem, see Walsh (1986), Theorem 2.6.

2.2.3 Further extensions of the stochastic integral

Our presentation has focused on the notion of infinite dimensional integral which is most relevant to this thesis. However, we would like to mention some other classical constructions and a few interesting developments from recent years, which are in a similar spirit.

The construction of the integrals for worthy martingale measures has been extended by Dalang (1999). In particular, this extension allows for treatment of spatially homogeneous noises, that is, the noises for which the correlation function is of the form $q(x, y) = f(x - y)$. We stress that this class of noises is not covered by Theorem 2.23 if

the domain of integration is $\mathbb{R}^d$. The extension of the integral is valid as long as f is a Fourier transform of a tempered non-negative measure μ on $\mathbb{R}^d$. Such measures satisfy

$$\int_{\mathbb{R}^d} \frac{1}{1+|\xi|^2}\mu(\mathrm{d}\xi) < \infty.$$

This condition is often referred to as the Dalang condition. In particular, this allows one to treat noises with covariance measures given by Riesz kernels (possibly multiplied by a smooth function ϕ), that is kernels of the form

$$q(x,y) = \phi(x)f(x-y) = \phi(x)(x-y)^{-\beta}, \quad \beta \in (0,2).$$

The main difference in comparison to the construction presented in Subsection 2.2.2.2 comes in the choice of the quantity used for bounds giving the dominating mesaure. Instead of the norm $\|\cdot\|_+$ we choose a norm $|\cdot|_0$ associated with the scalar product $\langle\cdot,\cdot\rangle_0$ (that is, $|g|_0 = \langle g,g\rangle_0^{1/2}$)given by

$$\langle f,g\rangle_0 = \left(\mathbb{E}\Big[\int_0^t \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(s,x,\cdot)g(s,y,\cdot)K_M(\mathrm{d}x\mathrm{d}y\mathrm{d}s)\Big]\right).$$

The rest of the construction is analogous. However, the result of the integration does not lead to a process which is another integrator. For details we refer to Dalang (1999), Section 2.

The integral with respect to a worthy martingale measure has been further extended by Dalang and Mueller (2003) to a function-valued stochastic integral. The main aim of this extension is to improve the theory for SPDEs which involve second order operators in the time derivative, such as the nonlinear wave equation.

An interesting extension, which forms a mixture of the two approaches discussed so far, has been proposed by Kurtz and Protter (1996). The definition of an integral with respect to a worthy martingale measure is extended to $H^{\#}$-semimartingales. We say that an adapted, $\mathbb{R}$-valued process Y is an $H^{\#}$-semimartingale if for any element h of the Hilbert space H, $Y(h,\cdot)$ is a semimartingale and Y is linear in the first variable. A good example of a process which does not fit the description of Subsection 2.13 but fits the setup of $H^{\#}$-semimartingales is the cylindrical Brownian motion (see Definition 2.15).

The theory of integration with respect to cylindrical martingales in topological vector spaces has also been developed, see Mikulevicius and Rozovskii (1998, 1999). The theory of stochastic integration with respect to cylindrical Lévy measures, simliar in the spirit to the theory presented in this chapter for the cylindrical Brownian motion, has also been developed recently, see, for example, Applebaum and Riedle (2010); Riedle (2014); Kumar and Riedle (2018).

An excellent comparison of the two classes of integrals introduced in this chapter, for spatially homogenous noises, can be found in Dalang and Quer-Sardanyons (2011).

2.3 Yamada-Watanabe Theorem

In this section we describe different notions of solution and different concepts of uniqueness for Stochastic Differential Equations (SDEs).

2.3.1 Yamada-Watanabe theorem for SDEs

In this section we consider the following stochastic differential equation

$$X_t = X(0) + \int_0^t b(t, X(s_-))\mathrm{d}s + \int_0^t \sigma(t, X(s_-))\mathrm{d}B_s, \tag{2.15}$$

where we assume that b, σ are Borel-measurable and $X(0)$ is independent of the filtration given by the Brownian motion B_t. We provide a description of different notions of solution and classical examples which illustrate the differences between them. Then we state the Yamada-Watanabe Theorem and its 'dual' result, the so called Engelbert's Theorem. The results presented in this section are classical, see, for example, Karatzas and Shreve (2012), Chapter 3 or Revuz and Yor (1999), Chapter IX .

All notions and results stated in this section are easily generalised to SDEs driven by a Brownian motion on $\mathbb{R}^d$. An overview of early results for more general integrators is given in Section 2.3.3.1.

2.3.1.1 Different notions of solution for SDEs

In this section we describe the various notions of solution to (2.15). We start with the most naive one.

Definition 2.33 (Solution to SDE)**.** *We call a pair of continuous (in time) stochastic processes* (X, B) *taking values in* $\mathbb{R}^2$ *a solution to equation* (2.15) *if the following two conditions are satisfied*

1. (X, B) *satisfies* (2.15)*,*

2. *The following inequality holds with probability one:*

$$\int_0^t (\|b(t, (X_{s_-}))\| + \|\sigma(t, (X_{s_-}))\|)\mathrm{d}s < \infty.$$

We note that this definition does not take into account any adaptability conditions for the underlying process. Taking them into account allow us to distinguish two completely different situations. In the first one, for any given Brownian motion B (input), we are able to construct a process X (output), which satisfies equation (2.15) and is adapted to the filtration associated with B. In the second one, given an equation of the form (2.15), we are able to construct a probability space, a process which is a Brownian motion B with respect to a filtration $\mathcal{F}_t$ on that probability space, and a process X measurable with respect to $\mathcal{F}_t$. We provide a rigorous definition for both of these notions of solution. We say that solutions of the first type are strong solutions, and solutions of the second type are weak solutions.

Definition 2.34 (Strong solution)**.** *Let* B_t *be a Brownian motion adapted to the filtration* $\mathcal{F}_t$*. A stochastic process* $\{X(t)\}_{t\geq 0}$ *is a strong solution of* (2.15) *with initial condition* $X(0)$ *and with respect to the Brownian motion* B *if*

1. (X, B) *solves* (2.15)*,*

2. X *is adapted to* $\mathcal{F}_t$*.*

Remark 2.35 (Alternative definition of strong solution)**.** *The definition of strong solution that we provided is just the most popular of many choices which lead to equivalent definitions in the context of equation* (2.15)*. For example, equivalently we could have required that*

1. *There exists a Borel-measurable map F such that*

$$X = F(B) \quad a.s.,$$

2. *X is adapted to $\mathcal{F}_t$.*

For the proof of equivalence of these two definitions we refer to Revuz and Yor (1999) Chapter IX, Theorem 1.7 and the remark stated after that Theorem.

Definition 2.36 (Weak solution)**.** *A weak solution to* (2.15) *is a triple* (X, B), $(\Omega, \mathcal{F}, \mathbb{P})$, $\mathcal{F}_t$ *which satisfies the following:*

1. (X, B) *solves* (2.15),

2. $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space with filtration, $\mathcal{F}_t$ which satisfies the usual conditions,*

3. B_t *and* $X(t)$ *are $\mathcal{F}_t$-adapted.*

2.3.1.2 Weak solutions and martingale problems

We would like to draw a connection between weak solutions and the martingale problem for the solution of the SDE. We begin by recalling the definition of the solution to the martingale problem.

Definition 2.37. *Let A be a generator of a process taking values in metric space E. We say that that a process X is a solution to a martingale problem for A if there exists a filtration $\mathcal{F}_t$ such that X_t is adapted to $\mathcal{F}_t$ and for any choice of f from $D(A)$*

$$f(X(t)) - f(X(0)) - \int_0^t Af(X(s))\mathrm{d}s$$

is a martingale with respect to the filtration $\mathcal{F}_t$. For any $\nu \in \mathcal{P}(E)$, we say that X is a solution to a martingale problem for (A, ν) if X is a solution to martingale problem for A and $X(0)$ has distribution ν.

Let A be a second order differential operator of the form

$$Af(x) = b(x)\partial_x f(x) + \frac{1}{2}\sigma^2(x)\partial_x\partial_x f(x).$$

Consider a weak solution to the SDE (2.15). Let $f \in C_c^2(\mathbb{R})$, the space of compactly supported, twice differentiable functions on $\mathbb{R}$. A simple application of Itô's formula yields

$$f(X(t)) - f(X(0)) - \int_0^t Af(X(s))\mathrm{d}s = \int_0^t \sigma(x)\partial_x f(X_t)\mathrm{d}B_s.$$

Therefore, since

$$\int_0^t \sigma(x)\partial_x f(X_t)\mathrm{d}B_s$$

is a local martingale, so is

$$f(X(t)) - f(X(0)) - \int_0^t Af(X(s))\mathrm{d}s. \tag{2.16}$$

This motivates the following definition:

Definition 2.38 (Martingale solution)**.** *We say that X is a martingale solution to* (2.15) *if the pair (X, ν), where $\nu = \mathbb{P} \circ X(0)^{-1}$, solves the martingale problem for A.*

The simple calculation which lead us to the definition of the martingale solution combined with a striking observation of Stroock and Varadhan (1972), results in the following theorem.

Theorem 2.39. *Every weak solution of* (2.15) *is a martingale solution. Every martingale solution gives rise to a weak solution. In other words, being a solution to a $(C[0,\infty))$ martingale problem given by* (2.16) *is equivalent to solving* (2.15) *in the weak sense.*

The usual proof of Theorem 2.39 involves an explicit construction of the driving noise B. However, this proof technique provides a serious challenge in the infinite dimensional setting. In Section 2.6, we present a different, more abstract, proof of the same statement in the SPDE case, adapted from the ideas of Kurtz (2010), who proves this result for infinite-dimensional SDEs. The crucial tool in this proof is the Markov Mapping Theorem. Theorem 2.39 can also be proved in an entirely analogous way.

2.3.1.3 Notions of uniqueness of solution for SDEs

In order to complete the question of well-posedness of our SDE, we need to define what we mean by uniqueness of solutions. In the case of strong solutions the notion of uniqueness is the most natural one - the solution is unique if two processes satisfying (2.15) (with the same initial condition) are indistinguishable.

Definition 2.40 (Strong uniqueness)**.** *Strong uniqueness holds for* (2.15) *with initial condition* $X(0)$*, if for any pair of solutions* $X_1(t)$*,* $X_2(t)$ *the paths of those solutions are indistinguishable, namely*

$$\mathbb{P}[X_1(t) = X_2(t) \quad \forall t \geq 0] = 1.$$

When we consider weak solutions, the picture is slightly more complicated - it is sensible to consider two different notions of uniqueness. Weak uniqueness, often referred to as uniqueness in law, requires that every two weak solutions (with the same initial condition) have the same finite-dimensional distributions.

Definition 2.41 (Weak uniqueness (uniqueness in law))**.** *Weak uniqueness holds for* (2.15) *with initial condition* $X(0)$*, if any pair of solutions* (X_1, B_2)*,* $(\Omega_1, \mathbb{P}_1, \mathcal{F}_1)$*,* $\mathcal{F}_t^1$ *and* (X_2, B_2)*,* $(\Omega_2, \mathbb{P}_2, \mathcal{F}_2)$*,* $\mathcal{F}_t^2$ *has the same law, namely for each finite collection of Borel sets* $\{\mathcal{B}_i\}_{i=1}^n$ *and any finite collection of times* $\{t_i\}_{i=1}^n$*, with* $t_i \geq 0$*,*

$$\mathbb{P}_1(X_1(t_1) \in \mathcal{B}_1, \ldots, X_1(t_n) \in \mathcal{B}_n) = \mathbb{P}_2(X_2(t_1) \in \mathcal{B}_1, \ldots, X_2(t_n) \in \mathcal{B}_n).$$

Remark 2.42. *As Theorem 2.39 does not assume uniqueness, and the uniqueness of the martingale problem is determined by the finite dimensional distributions, weak uniqueness is equivalent to uniqueness of the martingale problem related to the same SDE.*

It is also reasonable to ask whether weak solutions constructed from the same Brownian motion are the same. If this is the case, we say that pathwise uniqueness holds. Pathwise uniqueness requires that a pair of weak solutions with same initial data, driven by the same Brownian motion, are indistinguishable.

Definition 2.43 (Pathwise uniqueness)**.** *Pathwise uniqueness holds for* (2.15) *with initial condition* $X(0)$*, if any pair of solutions* (X_1, B)*,* $(\Omega, \mathbb{P}, \mathcal{F})$*,* $\mathcal{F}_t$*, and* (X_2, B)*,* $(\Omega, \mathbb{P}, \mathcal{F})$*,* $\mathcal{F}_t$ *is indistinguishable, namely,*

$$\mathbb{P}(X_1(t) = X_2(t) \quad \forall t \geq 0) = 1.$$

2.3.2 Illustrative examples

In this subsection we give examples to illustrate the different possibilities we might encounter, all are well known in the literature, see, for example, Revuz and Yor (1999), Chapter IX.

Notation 2.44. *By* $\operatorname{sgn}(x)$ *we mean a standard signum function, namely*

$$\operatorname{sgn}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

In the best possible scenario, the solution is strong and pathwise unique. A classical result on existence guarantees that this holds for a very broad class of functions.

Example 2.45. *Assume that functions* $b(t, x)$ *and* $\sigma(t, x)$ *are Lipschitz in* x *and grow at most linearly in* t*, namely that there exists a pair of constants* $C_1, C_2 > 0$ *such that*

$$|b(t, x)| + |\sigma(t, x)| \leq C_1(1 + |t|),$$
$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq C_2|x - y|.$$

Then equation (2.15) *has a unique strong solution.*

A classical example which shows a situation where a unique weak solution exists, but both strong existence and pathwise uniqueness fail is the so-called Tanaka example.

Example 2.46 (Tanaka equation)**.**

$$X_t = \int_0^t \operatorname{sgn}(X(s)) \mathrm{d}B_s.$$

A straightforward modification of Tanaka's example leads us to an equation for which strong existence (and therefore weak existence) holds, but there is no uniqueness.

Example 2.47.

$$X_t = \int_0^t \mathbf{1}_{\{X(s)\neq 0\}} \mathrm{d}B_s.$$

Finally, another modification of Tanaka's example leads us to an equation for which only weak existence can be proved.

Example 2.48.

$$X_t = X_0 + \int_0^t \mathbf{1}_{\{X(s)\neq 0\}} \operatorname{sgn}(X(s)) \mathrm{d}B_s.$$

Remark 2.49. *A large class of examples for which either strong or weak uniqueness fails, even for continuous σ and b, was given by Barlow (1982).*

At first sight it seems that this list of examples is rather incomplete. However, the results presented in the next section guarantee that for the SDE (2.15), this list is exhaustive.

2.3.3 Classical Yamada-Watanabe Theorem

The Yamada-Watanabe Theorem, first proved for SDEs driven by Brownian motion by Yamada and Watanabe (1971), explains the relations between different concepts of uniqueness and existence.

Theorem 2.50 (Yamada-Watanabe Theorem)**.** *If a weak solution to* (2.15) *exists and is pathwise unique, there exists a strong solution to* (2.15)*. What is more, weak uniqueness holds.*

The following theorem, first proved in the context of SDEs driven by Brownian motion by Gihman and Skorohod (1979), shows that a 'dual' of this statement is also true.

Theorem 2.51 (Engelbert's Theorem)**.** *If a strong solution to* (2.15) *exists and if weak solutions to* (2.15) *are weakly unique, than all weak solutions of* (2.15) *are pathwise unique.*

This theorem is known as Engelbert's Theorem due to its useful generalisation by Engelbert (1991), which will be discussed in the next section.

We conclude this subsection with a result by Cherny (2002), which gives a full classification of the possible interplay between types of existence and types of uniqueness. This result states that for the SDEs driven by Brownian motion, the list of examples in Section 2.3.2 is exhaustive, namely that there are only four possible configurations of existence and uniqueness:

1. there exists a unique, strong solution, which implies weak existence and uniqueness, see Example 2.45;

2. there exists a non-unique strong solution, which implies weak existence, but this solution is not unique in either the weak or the pathwise sense, see Example 2.46;

3. there exists a unique weak solution, but strong existence and pathwise uniqueness fails, see Example 2.47;

4. there exists a weak solution, but the strong existence and every type of uniqueness fails, see Example 2.48.

In the light of this result it may seem that the notion of the pathwise uniqueness is not very useful. However, as argued before, this property might be easier to verify, and the importance of the Yamada-Watanabe Theorem lies in the fact that just weak existence and pathwise uniqueness guarantee all possible forms of uniqueness.

2.3.3.1 An overview of Yamada-Watanabe results for more general SDEs

So far our discussion has focused on a simple SDE driven by a Brownian motion. However, one would like to extend the results to other SDEs driven by more complicated integrators. In this section we point towards several results in that direction.

Here, we are interested in equations of the form

$$X_t = X(0) + \int_0^t \sigma(s, X_{s_-}) \mathrm{d}Z_s, \tag{2.17}$$

where Z_s is a semimartingale. To state a first result, due to Jacod (1980), we need one more definition.

Definition 2.52 (Joint uniqueness in law)**.** *We say that a solution to equation* (2.17) *is jointly unique in law if for any pair of solutions* (X_1, Z_1) *and* (X_2, Z_2)

$$(X_1(t), Z_1(t); t \geq 0) \stackrel{d}{=} (X_2(t), Z_2(t); t \geq 0).$$

This definition is not, in general, equivalent to uniqueness in law of Definition 2.41, see Brossard (2003). We may now state the Yamada-Watanabe Theorem for equations of the form (2.17) and an analogue of Theorem 2.51.

Theorem 2.53 (Yamada-Watanabe Theorem for SDEs driven by semimartingales)**.** *If a weak solution to* (2.17) *exists and is pathwise unique, than it is jointly unique in law and strong.*

Theorem 2.54 (Engelbert's Theorem, Engelbert (1991), Theorem 2)**.** *If a strong solution to* (2.17) *exists, and if the solution is jointly unique in law, every weak solution is pathwise unique and therefore the solutions are strong and unique.*

Remark 2.55. *The careful reader might have observed that the statements of Theorems 2.50, 2.51 and Theorems 2.53, 2.54 differ only in the type of uniqueness in law that can be deduced. This difference is nicely explained by Theorem 3.1 of Cherny (2002), which states that for SDEs driven by Brownian motion (possibly multidimensional) joint uniqueness in law is equivalent to uniqueness in law.*

The full analogue of the Yamada-Watanabe Theorem 2.50 (with uniqueness in law instead of joint uniqueness in law) is available for one-dimensional SDEs with jumps, see Barczy et al. (2014).

2.4 Solutions of general stochastic models

In this section a much more general approach to stochastic models is presented. We treat models that might not necessarily be equations, but are composed from a set of constraints, e.g. equations or inequalities. We also do not put any restrictions on the dimensionality of the problems. In particular, this approach, which closely follows Kurtz (2007, 2014), allows for treatment of SPDEs. The main aim of this subsection is to

introduce the framework in which much more general versions of the Yamada-Watanabe and Engelbert Theorems can be proved. As the terminology used throughout this section is very different from the standard one, we discuss it in Subsection 2.4.1. We state the General Yamada-Watanabe Theorem 2.66 at the end of this subsection. We discuss the notions of solution to SPDE in Subsection 2.4.2. Finally, we prove the Yamada-Watanbe Theorem for SPDEs in Subsection 2.4.3.

2.4.1 Framework and General Yamada-Watanabe Theorem

Consider a stochastic model composed of two random variables, Y (inputs) and X (outputs), which are related by some constraints Γ. These constraints could be, for example equations or inequalities. We assume that the random variables take values in the complete, separable metric spaces S_1 and S_2.

Consider a space of probability measures on $S_1 \times S_2$, $\mathcal{P}(S_1 \times S_2)$. For two random variables X and Y, taking values in S_1 and S_2 respectively, let $\mu_{X,Y} \in \mathcal{P}(S_1 \times S_2)$ denote their joint probability distribution. Let ν be the marginal distribution of $\mathcal{P}(S_1 \times S_2)$.

Now our stochastic model is specified by a pair (Γ, ν). Let $\mathcal{S}_{\Gamma,\nu} \subset \mathcal{P}(S_1 \times S_2)$ be a set of joint solution measures, that is $\mu_{X,Y} \in S_{\Gamma,\nu}$ if (X, Y) is a solution to (Γ, ν).

To demystify the rather abstract setup presented here, let us return to the example of a simple one-dimensional SDE considered in Section 2.3.1.

Example 2.56 (SDE driven by Brownian motion in the general framework)**.** *Consider equation* (2.15)*. For this equation, the stochastic process which determines the inputs is just a Brownian motion B_t, which takes values in $S_1 = C[0, \infty)$. The output process, X, is valued in $S_2 = (\mathbb{R} \times C[0, \infty))$, (where the first component represent the initial condition). Both S_1 and S_2 are complete, separable metric spaces. The marginal ν is given by the measure characterising the Brownian motion. Every weak solution gives an example of a joint solution measure. However, it is not* a priori *true that every element of $\mathcal{S}_{\Gamma,\nu}$ is a weak solution, since the adaptability is not a part of our set of constraints.*

Let us rephrase some of the definitions of Subsection 2.3.1 in our new framework.

Definition 2.57 (Notions of uniqueness of solutions, Kurtz (2014), Definition 1.4)**.** *Consider a stochastic model of the form (Γ, ν).*

1. *We say that pathwise (pointwise) uniqueness holds if for any pair of processes X_1, X_2 defined on the same probability space and an input process Y such that $\mu_{X_1,Y}, \mu_{X_2,Y} \in \mathcal{S}_{\Gamma,\nu}$, almost surely $X_1 = X_2$.*

2. *We say that joint uniqueness in law holds if $\mathcal{S}_{\Gamma,\nu}$ contains at most one measure.*

3. *We say that uniqueness in law holds if all measures $\mu \in \mathcal{S}_{\Gamma,\nu}$ have the same marginal distribution on S_1.*

Remark 2.58. *Observe that it is immediately clear from Definition 2.57 that joint uniqueness in law implies uniqueness in law. An example for the converse being false is once again given by Brossard (2003).*

Definition 2.59 (Strong solution in the sense of Kurtz)**.** *We say that a solution (X, Y) of (Γ, ν) is strong if there exists a Borel measurable function such that*

$$X = F(Y).$$

The following theorem relates two of the notions of solution.

Theorem 2.60 (Kurtz (2014), Theorem 1.5)**.** *If a joint solution measure exists and is pathwise unique, the solution is jointly unique in law.*

We observe that the notions of solution defined in Definition 2.57 and Definition 2.59 are significantly different from those in Subsection 2.3.1.3, as in Subsection 2.3.1.3 we required the solution to be adapted to the filtration of the input process. In order to make the connection between these two regimes we follow Kurtz (2007, 2014) and define the notion of compatibility.

Definition 2.61 (Compatibility structure)**.** *Let $\mathcal{A}$ be a set of indices. For each $\alpha \in \mathcal{A}$, let $\mathcal{B}_\alpha^{S_i}$ be a sub-algebra of all Borel sets on S_i. We say that the collection of sets*

$$\mathcal{C} = \bigcup_{\alpha \in \mathcal{A}} \left\{ \mathcal{B}_\alpha^{S_1}, \mathcal{B}_\alpha^{S_2} \right\}$$

is a compatibility structure.

Definition 2.62 (Compatible solutions)**.** *Let X and Y be a pair of random variables with values in S_1 and S_2 respectively and with complete filtrations $\mathcal{F}_t^X$, $\mathcal{F}_t^Y$ respectively.*

We say that X and Y are $\mathcal{C}$-compatible if for each index $\alpha \in \mathcal{A}$ and $h \in L^1$

$$\mathbb{E}[h(Y)|\mathcal{F}_\alpha^X \vee \mathcal{F}_\alpha^Y] = \mathbb{E}[h(Y)|\mathcal{F}_\alpha^Y]$$

We will denote a set of $\mathcal{C}$-compatible solutions by $\mathcal{S}_{\Gamma,\mathcal{C},\nu}$.

Remark 2.63 (Temporal compatibility)**.** *If $\mathcal{A} = [0,\infty)$, and both $\mathcal{F}_t^X$ and $\mathcal{F}_t^Y$ are complete filtrations, the compatibility conditions reduce to the temporal compatibility, which was already present in Jacod (1980). In this case*

$$\mathcal{F}_\alpha^X \vee \mathcal{F}_\alpha^Y = \mathcal{F}_t^{X,Y},$$

and the compatibility criterion reduces to

$$\mathbb{E}[h(Y)|\mathcal{F}_t^{X,Y}] = \mathbb{E}[h(Y)|\mathcal{F}_t^Y],$$

which should hold for each $h \in L^1$.

Definition 2.64 (Partially compatible solutions)**.** *Let $\mathcal{H} \subset L^1$. We say that two random variables are partially-$\mathcal{C}$-$\mathcal{H}$ compatible if for each $h \in \mathcal{H}$*

$$\mathbb{E}[h(Y)|\mathcal{F}_\alpha^X \vee \mathcal{F}_\alpha^Y] = \mathbb{E}[h(Y)|\mathcal{F}_\alpha^Y].$$

We will denote a set of partially $\mathcal{C},\mathcal{H}$-compatible solutions by $\mathcal{S}_{\Gamma,(\mathcal{C},\mathcal{H}),\nu}$.

The major difficulty in applying Kurtz's version of the Yamada-Watanabe Theorem lies in the fact that it is not immediately clear whether the joint solution measures which appear in the statements of the theorems correspond to the usual definitions of weak solutions, as it is usually assumed that the solution is adapted to the filtration of the input process. However, the following theorem from Kurtz (2014) clarifies that the compatibility criterion guarantees the required adaptedness, at least for strong solutions.

Theorem 2.65 (Kurtz (2014), Proposition 2.13)**.** *Let X be a strong, compatible solution, then $\mathcal{F}_\alpha^X \subset \mathcal{F}_\alpha^Y$ for each $\alpha \in \mathcal{A}$. In particular, if X is temporally compatible, it is adapted*

to $\mathcal{F}_t^Y$. *Conversely, if* $\mathcal{F}_\alpha^X \subset \mathcal{F}_\alpha^Y$ *for each* $\alpha \in \mathcal{A}$ *and* $\sigma(X) \subset \vee_{\alpha\in\mathcal{A}} F_\alpha^X$ *then* X *is a strong, compatible solution.*

We are now ready to state the general Yamada-Watanabe Theorem Kurtz (2007, 2014). This result does not require any specific structure of the stochastic model; instead the constraints are applied via compatibility and partial compatibility conditions.

Theorem 2.66 (Kurtz (2014), Theorem 3.4)**.** *Suppose that the solution* X*, jointly partially compatible with* Y *is pathwise unique, that is, whenever* X_1 *and* X_2 *are jointly partially compatible with* Y*, and* $\mu_{X_1,Y}, \mu_{X_2,Y} \in \mathcal{S}_{\Gamma,(\mathcal{C},\mathcal{H}),\nu}$ *then* $X_1 = X_2$*. In addition, suppose that there exists a compatible solution, so that* $\mathcal{S}_{\Gamma,\mathcal{C},\nu} \neq \emptyset$*. Then there exists a unique partially compatible solution and it is compatible.*

2.4.2 Notions of solutions for SPDE

In this subsection we use the theory of stochastic integration from Section 2.2 and ideas from classical deterministic theory to specify what we mean by a solution to equations (2.1) and (2.2). We begin with the concept of mild solutions.We then discuss analytical solutions, clearly better suited to the Hilbert space formulation. Finally, we discuss the probabilistic notion of weak and strong solution. The terminology, and some of the statements, are borrowed from Prévôt and Röckner (2007).

The mild solution is heavily based on semigroup theory. We represent the solution using the variation of constants formula, first introduced in the context of infinite dimensional analysis by Ball (1977).

Definition 2.67 (Mild solution)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a complete probability space and let* $\mathcal{F}_t$ *be a complete, right continuous filtration. Let* $p_t(x)$ *be the standard heat kernel, as defined in* (2.4)*. Let* M *be a martingale measure. A mild solution to* (2.1)*, if it exists, is an* $\mathcal{F}_t$*-adapted, jointly measurable process* $u(t,x)$ *such that*

$$u(t,x) = \int_{\mathbb{R}^d} p_t(x-y)u(0,x)\mathrm{d}x + \int_0^t \int_{\mathbb{R}^d} p_{t-s}(x-y)b(s,y,u(s,y))\mathrm{d}y\mathrm{d}s + \int_0^t \int_{\mathbb{R}^d} p_{t-s}(x-y)\sigma(s,y,u(s,y))M(\mathrm{d}t,\mathrm{d}y), \quad (2.18)$$

where the integral with respect to M *is interpreted in the way defined in Subsection 2.2.2.2.*

Our definition of mild solution puts some conditions on the regularity of solutions. We discuss these in Section 2.5.

In order to formulate an analogue of Definition 2.67 for the Hilbert space formulation we define the heat semigroup as

$$S_t\Phi = \int_{\mathbb{R}^d} p_t(x,y)\Phi(y)\mathrm{d}y$$

whenever the integral makes sense.

Definition 2.68 (Mild solution, Hilbert space approach)**.** *Let H be a Hilbert space. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and let $\mathcal{F}_t$ be a complete, right continuous filtration. Let W be a Q-Wiener process or a cylindrical Brownanian motion. An H-valued, predictable process $X(t)$ is a mild solution to equation* (2.2) *with initial condition $X(0)$, if, for each $t > 0$,*

$$X(t) = S_tX(0) + \int_0^t S_{t-s}B(s,X(s))\mathrm{d}s + \int_0^t S_{t-s}\sigma(s,X(s))\mathrm{d}W(s), \tag{2.19}$$

where the integral with respect to W is interpreted in the way defined in the Subsection 2.2.2.1.

Let W be a $Q-Wiener$ measure. Since by equation (2.12) it is possible to construct a martingale measure from W, the Definitions 2.67, 2.68 are equivalent for equations driven by Q-Wiener measures and a wide class of coloured noises. Using the argument exploited in our construction of cylindrical Brownian motion we could extend this equivalence to the white noise case, at least in spatial dimension one.

Analytically strong and analytically weak solutions bear the closest resemblance to those present in the literature for deterministic partial differential equations. We use the word 'analytical' (even though this terminology is not widespread) to distinguish them from the probabilistic notions of weak and strong solutions. For simplicity, we formulate these for equations driven by Q-Wiener processes and cylindrical Brownian motions.

Definition 2.69 (Analytically strong solution)**.** *We say that a predictable H-valued process $X(t)$ with initial condition $X(0)$ is an analytically strong solution to equation*

(2.2) *if*

$$X(t) = X(0) + \int_0^t [AX(s) + B(s, X(s))]\mathrm{d}s + \int_0^t \sigma(s, X(s))\mathrm{d}W(s),$$

where the integral with respect to W is interpreted in the way defined in Subsection 2.2.2.1

Definition 2.70 (Analytically weak solutions)**.** *We say that a predictable H-valued process $X(t)$ with initial condition $X(0)$ is an analytically weak solution to equation* (2.2) *if for every test function f in the domain $\mathcal{D}(A^*)$ of the adjoint of A,*

$$\langle X(t), f\rangle = \langle X(0), f\rangle + \int_0^t [\langle X(s), A^* f\rangle + \langle B(s, X(s))f\rangle]\mathrm{d}s + \int_0^t \langle \sigma(s, X(s))f\rangle \mathrm{d}W(s).$$

Obviously each analytically strong solution is also analytically weak, but not the other way round. The analytically weak formulation of equation (2.1) is given by

$$\begin{aligned}\int_{\mathbb{R}^d} u(t,x) = &\int_{\mathbb{R}^d} \phi(x)u(0,x)\mathrm{d}x + \int_0^t \int_{\mathbb{R}^d} u(s,x)\Delta\phi(x)\mathrm{d}x \\ &+ \int_0^t \int_{\mathbb{R}^d} \phi(x)b(s,y,u(s,x))\mathrm{d}x\mathrm{d}s + \int_0^t \int_{\mathbb{R}^d} \phi(x)\sigma(s,y,u(s,x))W(\mathrm{d}t,\mathrm{d}x), \quad (2.20)\end{aligned}$$

since the Laplace operator is self-adjoint.

The definitions of (stochastically) weak and strong solutions of SPDEs are completely analogous to their SDE counterparts discussed in Subsection 2.3.1.1.

Definition 2.71 (Stochastically weak solution)**.** *We say that a stochastically weak solution to SHE exists if one can find a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a right continuous filtration $\mathcal{F}_t$ such that*

1. *almost surely the following is satisfied*

$$\int_0^t |b(s, X(s))|\mathrm{d}s + \int_0^t \|\sigma(s, X(s))\|^2 \mathrm{d}s < \infty.$$

2. *An analytically weak solution to SHE* (2.2) *exists.*

Definition 2.72 (Stochastically strong solution)**.** *We say that a stochastically weak solution to SHE* (2.2) *is stochastically strong if for any given random noise supported*

on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *with complete, right continuous filtration* $\mathcal{F}_t$*, the solution is* $\mathcal{F}_t$*-adapted.*

To discuss uniqueness of solutions, we need to specify the state space of the process. In what follows, $\mathcal{H}$ denotes a class of functions (e.g. continuous, smooth, càdlàg).

Definition 2.73 (Weak uniqueness)**.** *We say that the weak uniqueness holds for the solution of equation* (2.2) *if for every pair of weak solutions* $X_1, X_2 \in \mathcal{H}$ *their finite dimensional distributions are equal.*

Definition 2.74 (Pathwise uniqueness)**.** *Pathwise uniqueness holds for the solution of equation* (2.2) *if a pair of weak solutions* $X_1, X_2 \in \mathcal{H}$ *is driven by the same noise, then* X_1, X_2 *are indistinguishable.*

Definition 2.75 (Strong uniqueness)**.** *We say that strong uniqueness holds if every pair of strong solutions* $X_1, X_2 \in \mathcal{H}$*, driven by a noise defined on a common probability space is indistinguishable.*

We notice that the difference between strong and pathwise uniqueness is very subtle - in the case of pathwise uniqueness we do not know whether the solution does exist for any probability space (as we are allowed to choose the filtration, the probability space and the noise), whereas in the case of strong uniqueness we require that the property holds for every probability space. We also observe that our definitions of stochastically weak and strong solutions imply that if the stochastically strong solution is unique, this solution is also the unique stochastically weak solution.

Remark 2.76. *We have formulated the notions of solution and uniqueness for the SHE. If one considers a different equation, the definitions of analytical solutions are completely analogous. A definition analogous to Definition 2.68 is valid when we replace* S_t *with any* C_0 *semigroup of linear operators. For more details we refer to Da Prato and Zabczyk (2008), Appendix A. For a more analytical exposition of semigroup theory we refer to Lunardi (2012, 2018). We also point out that for fully nonlinear equations (for example those in which instead of the Laplacian ones consider a p-Laplacian, that is, for* $1 < p < \infty$*,* $\Delta_p u := \nabla \cdot (|\nabla u|^{p-2} \nabla u)$*) many results have been established in the regime of variational solutions, see Prévôt and Röckner (2007), Liu and Röckner (2015).*

2.4.3 General Yamada-Watanabe Theorem and SPDEs

In this subsection we focus on applications of the General Yamada Watanabe Theorem 2.79 to SPDEs. We would like to point out that our approach, although it seems to be the most robust, is not the only one available. In the variational framework, the result has been proved by Röckner et al. (2008) for evolution equations and by Tappe (2013) for semilinear SPDEs with path-dependent coefficients. In the setting of mild solutions, the Yamada-Watanabe Theorem has been treated in impressive detail by Ondreját (2004).

The following lemma is the crucial tool in the application of Theorem 2.66 in the context of SPDEs. The proof mimics Athreya et al. (2018), Lemma 5.7, which in turn, for the crucial part, mimics the proof of Kurtz (2014), Lemma 2.4. For simplicity, we prove the lemma under the assumptions of Theorem 2.81. However, the generalisations of the proof are straightforward. A similar, but different statement appeared in Lemma 5.1 from Rippl (2012).

We shall need one more definition.

Definition 2.77 (Exponentially weighted L^p space)**.** *Fix $\lambda > 0$ and $1 \leq p \leq \infty$. We say that a function $f : \mathbb{R}^d \to \mathbb{R}$ belongs to $L^p_\lambda(\mathbb{R}^d)$ if*

1. *for $1 \leq p < \infty$*

$$\|f\|^p_{L^p_\lambda} = \int_{\mathbb{R}^d} |f(x)|^p e^{-\lambda|x|} \mathrm{d}x < \infty,$$

2. *for $p = \infty$*

$$\|f\|_{L^\infty_\lambda} = \sup_{x \in \mathbb{R}^d} |f(x)| e^{-\lambda|x|} < \infty.$$

The space L^p_λ is a Banach space with the norm $\|\cdot\|_{L^p_\lambda}$ and the corresponding topology.

Lemma 2.78. *If the solution to the SHE* (2.1) *is weak, it is compatible.*

Proof Let W be either coloured or white noise. Let (X, W) be a weak solution to equation (2.1) adapted to a filtration $\mathcal{F}_t$. We now formulate the problem in the terminology of Section 2.4. Define $S_1 = C\Big([0,\infty], \mathcal{S}'(\mathbb{R}^d)\Big)$ and $S_2 = C\Big([0,\infty], L^\infty_\lambda(\mathbb{R}^d)\Big)$.

We observe that our stochastic input, namely the noise W, can be defined on S_1 and by Theorem 2.81, S_2 is the solution space for SHE. Let $\mathcal{P}(S_1 \times S_2)$ be the space of probability probability measures on $(S_1 \times S_2)$. For a random variable X taking values in S_1 and a random variable Y taking values in S_2, let $\mu_{X\times Y}$ be the joint probability distribution of (X, Y). The set of constraints Γ, is given by the SHE (2.1). We want to show that every weak solution of SHE (2.1) is compatible in the sense of Definition 2.62.

Define $Y(s) = (X(0), W(s))$. Consider a pair of processes $Y^{\leq t}, Y^{\geq t}$ defined for $s \in [0, T]$ as

$$Y^{\leq t}(s) = (W_{s\wedge t}) \qquad Y^{\geq t}(s) = (W_{t+s} - W_t).$$

Obviously, $\mathcal{F}_t^W \subset \mathcal{F}_t^X \vee \mathcal{F}_t^W \subset \mathcal{F}_t$.Now we may observe that the processes $Y^{\leq t}$ is measurable with respect to $\mathcal{F}_t^W$. Since W is the white (or colored) noise with respect to filtration $\mathcal{F}_t$, it is also white (or colored) with respect to $\mathcal{F}_t^X \vee \mathcal{F}_t^W$. Therefore $Y^{\geq t}$ is independent of $\mathcal{F}_t^X \vee \mathcal{F}_t^W$.

Let Σ be a distribution of W on S_1. Now we observe that, by continuity of W, for each $h \in L^1(S_2, \Sigma)$ there exists a measurable function h_t defined on $S_1 \times S_2$, such that

$$h(W) = h_t(Y^{\leq t}, Y^{\geq t})$$

We may now directly mimic the calculation from the proof of Kurtz (2014), Lemma 2.4.

$$\begin{aligned}\mathbb{E}\Big[h(Y)|\mathcal{F}_t^X \vee \mathcal{F}_t^W\Big] = \mathbb{E}\Big[h_t(Y^{\leq t}, Y^{\geq t})|\mathcal{F}_t^X \vee \mathcal{F}_t^W\Big] = \\ \mathbb{E}\Big[\int h_t(Y^{\leq t}, y)\mu_{Y^{\geq t}}(\mathrm{d}y)|\mathcal{F}_t^X \wedge \mathcal{F}_t^W\Big] = \int h_t(Y^{\leq t}, y)\mu_{Y^{\geq t}}(\mathrm{d}y) \\ = \mathbb{E}\Big[\int h_t(Y^{\leq t}, y)\mu_{Y^{\geq t}}(\mathrm{d}y)|\mathcal{F}_t^W\Big] = \mathbb{E}\Big[h(Y)|\mathcal{F}_t^W\Big],\end{aligned}$$

where we have used tower property several times. Therefore, by Definition 2.62, Y is compatible. □

Theorem 2.79 (Yamada-Watanabe Theorem for SPDE)**.** *Assume that a (stochastically) weak solution to equation* (2.2) *exists. Assume that the weak solutions to* (2.2) *are pathwise unique. Then the only solution to* (2.2) *is a unique strong solution.*

Proof The statement of Theorem 2.79 follows immediately from Theorem 2.66 and an application of Lemma 2.78, as the weak solutions are compatible, and therefore partially compatible. What is more, the pointwise uniqueness of Theorem 2.66 is, in our context, just pathwise uniqueness. □

Remark 2.80. *Observe that in the proof of Lemma 2.78 we only used the properties of the noise, not the structure of the equation. Therefore, as long as the notion of solution is well defined and the noise has independent increments, the proof of the Lemma 2.78 could remain unchanged and the conclusions of Lemma 2.78, and therefore Theorem 2.79 apply. This flexibility explains the relative complexity of our proof of Lemma 2.78.*

2.5 Stochastic heat equation

In this section we apply the constructions of the noise and integral and various notions of solution to give a meaning to SHE(2.1). In order to prove existence of solutions, we need to impose some restrictions on the regularity of the coefficients. These conditions are easily satisfied for our applications in later parts of the thesis. We also recall a pathwise uniqueness result which will be useful in the later part of the thesis.

We begin by clarifying the regularity of solutions. In what follows it is convenient to use the notation $\lesssim$ to denote an inequality which holds up to a constant. The following existence result can be found in Mytnik et al. (2006), Theorem 1.2 and Theorem 1.8; Rippl and Sturm (2013), Theorem 1.1.

Theorem 2.81 (Existence of solutions for SHE)**.** *Consider the SHE* (2.1) *on* $\mathbb{R}^d$ *driven by coloured noise with correlation kernel* $q(x, y)$*, bounded by a Riesz potential, that is*

$$q(x, y) \lesssim (1 + |x - y|^{\alpha}), \quad \alpha \in (0, 2 \wedge d). \tag{2.21}$$

Let the initial condition u_0 *belong to the weighted space* L^{∞}_{λ}*. In addition, assume the following set of conditions on the regularity of the coefficients of equation* (2.1)*:*

1. *the functions* b *and* σ *grow at most linearly at infinity, that is*

$$|\sigma(t, x, u)| + |b(t, x, u)| \lesssim c_2(1 + |u|), \tag{2.22}$$

2. the function σ satisfies

$$|\sigma(t,x,u_1)-\sigma(t,x,u_2)| \lesssim e^{A_1|x|}(1+|u_1|+|u_2|)^{A_2}|u_1-u_2|^{\gamma}, \tag{2.23}$$

where $\gamma \in (0,1)$, $c_3, A_1, A_2 > 0$,

3. the function b is Lipschitz, that is

$$|b(t,x,u_1)-b(t,x,u_2)| \lesssim |u_1-u_2|. \tag{2.24}$$

Then there exists a solution to (2.1)*. This solution is continuous in time, and for any $T, p, \lambda > 0$, it satisfies the following bound:*

$$\mathbb{E}\left[\sup_{0\le t\le T}\sup_{x\in R^d}|u(t,x)|^p e^{-\lambda|x|}\right] < \infty.$$

Since for some of our proofs later on, especially the proof of Theorem 2.85, it is more convenient to think of the analytically weak solutions, the following results turns out to be useful.

Remark 2.82. *Under assumptions of Theorem 2.81 the mild and analytically weak formulations of* (2.1) *are equivalent, see discussion on p. 1917 ofMytnik et al. (2006).*

The following result guarantees pathwise uniqueness for the SHE with coloured noise.

Theorem 2.83 (Pathwise uniqueness for SHE, Rippl and Sturm (2013), Theorem 1.2)**.** *Under the assumptions of Theorem 2.81, equation* (2.1) *with initial data in L^{∞}_{λ} has a unique pathwise solution whenever $\alpha < 2(2\gamma-1)$.*

An application of our Yamada-Watanabe Theorem, in combination with the weak existence result and Theorem 2.83, guarantees that the only solutions to SHE (2.1) with coloured noise satisfying the assumptions of Theorem 2.81 are strong and unique. In particular, weak uniqueness holds. We record this observation as a corollary.

Corollary 2.84. *Under the assumptions of Theorem 2.83, SHE* (2.1) *has a unique strong solution. In particular, weak uniqueness holds.*

2.6 Equivalence of the martingale problem and SPDE

In this section we present a result that, under mild conditions, guarantees equivalence between the solutions of a martingale problem and solutions of the related SPDE. This result is an analogue of Theorem 2.39. We present a proof for the SHE, however, our proof requires only minor modifications to be adapted to a wide class of SPDEs.

Theorem 2.85. *Write w for a continuous representative of the density of the $\mathcal{M}_\lambda$-valued process. Let b, σ, g be functions satisfying the conditions of Theorem 2.81. Suppose that for all $f \in C_c^\infty(\mathbb{R}^d)$,*

$$\langle w_t, f\rangle - \langle w_0, f\rangle - \int_0^t \left\{\frac{1}{2}\langle w, \Delta f\rangle + \langle b(s,w), f\rangle\right\}\mathrm{d}s \tag{2.25}$$

is a martingale with quadratic variation

$$\int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} g(x,y)\sigma(s,x,w))\sigma(s,y,w))f(x)f(y)\mathrm{d}x\mathrm{d}y\mathrm{d}s.$$

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

$$dw = \left\{\frac{1}{2}\Delta w + b(t,x,w)\right\}dt + \sigma(t,x,w)W(\mathrm{d}t,\mathrm{d}x), \tag{2.26}$$

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 2.85 is a simple application of Itô's formula.

We closely follow Kurtz (2010), 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). This proof can be adapted further without any major difficulties to cover a wider class of equations, i.e. more singular coefficients b, σ in (2.26) or more general parabolic or hyperbolic equations, all under less restrictive conditions on the regularity of the density

w. For convenience, we first recall the Markov Mapping Theorem in the form stated there.

Theorem 2.86 (Markov Mapping Theorem, Kurtz (2010)), 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, $\gamma : E \to E_0$ be Borel measurable, and α be a transition function from E_0 into E ($y \in E_0 \to \alpha(y, \cdot) \in \mathcal{P}(E)$ 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 $M_{E_0}[0, \infty)$ (space of measurable functions mapping $[0, \infty)$ to E_0, with topology given by convergence in Lebesgue measure) as $U = \gamma \circ V$ and

$$\mathbb{P}\left[V(t) \in \Gamma | \widehat{\mathcal{F}}_t^U \right] = \alpha(U(t), \Gamma), \quad \Gamma \in \mathcal{B}(E), t \in T^U, \tag{2.27}$$

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

The proof of Theorem 2.85 is based on the introduction of an auxiliary process Z, whose generator plays the role of B in the statement of Theorem 2.86. 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 X. 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 a Hilbert space $W^{-1-d,2}(\mathbb{R}^d)$ (see, i.e. Rippl (2012)), that is on a dual of a 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

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

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

$$Y_i = Y_i(0) + \Big(W_t(\phi_i) \mod 2\pi \Big).$$

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

$$W_t(\phi_i) = \int_0^t (-\sin(Y_i(s)), \cos(Y_i(s)))d\beta_i(s). \tag{2.28}$$

Remark 2.87. *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

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

where M is a solution of the martingale problem (2.25) and $Y = \{Y_i\}_{i \geq 1}$,

Let A be the generator associated with the martingale problem (2.25). Its domain is given by

$$\mathcal{D}(A) = \Big\{ \psi_{F,f} = F(\langle \cdot, w \rangle) : f \in C_c^\infty(\mathbb{R}^d), F \in C^1(\mathbb{R}) \Big\}. \tag{2.29}$$

Let $\widehat{A}$ be the generator of the process Z. Let $D_0([0, 2\pi))$ be the collection of functions for which

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

The domain of $\widehat{A}$ is given by

$$D(\widehat{A}) = \left\{ f_0(x) \prod_{i=1}^{\infty} f_i(y_i) : f_0 \in \mathcal{D}(\mathcal{A}), f_i \in D_0 \right\}.$$

An application of Itô's formula guarantees that $\widehat{A}$ can be written in the form

$$\widehat{A}f = L_x f(x, y) + L_y f + L_{xy} f(x, y),$$

where $L_x f = Af$, L_y is the generator of Y and

$$L_{xy} f = \sum_{i,k=1}^{\infty} c_{ik} \partial_{x_i} \partial_{y_i} f$$

and the constants c_{ik} are determined by the Meyer processes $[M, Y_i]$. Their exact value won't concern us. To show that solutions to a martingale problem for $\widehat{A}$ are weak solutions to (2.26), we need a following Lemma, which we once again borrow from Kurtz (2010).

Lemma 2.88 (Kurtz (2010), Lemma A.1)**.** *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*

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

Suppose $\mathcal{D}(A)$ is an algebra and that $f \circ X$ is càdlàg for each $f \in \mathcal{D}(A)$. Let $f_1 \cdots, f_m \in \mathcal{D}(A)$ and $h_1, \cdots, h_m \in B(E)$. Then

$$M(t) = \sum_i \int_0^t g(X_{s_-}) \mathrm{d}M_{f_i}(s)$$

is a square integrable martingale with Meyer process

$$\langle M\rangle_t = \sum_{i,j}\int_0^t g_i(X_s)g_j(X(s)) \times \Big[Af_if_j(X(s)) - f_i(X(s))Af_j(X_s) - f_j(X_s)Af_i(X_s)\Big]\mathrm{d}s \quad (2.30)$$

Lemma 2.89. *Every solution of the martingale problem defined by $\widehat{A}$ is a solution of* (2.26)*, with W defined by* (2.28)*.*

Proof The statement follows by application of Lemma 2.88 with $f_1(x,y) = f(x)$, $f_2(x,y) = \cos(y_i)$, $f_3(x,y) = \sin(y_i)$, $h_1(x,y) = 1$, $h_2(x,y) = f'(x)\sin(y_i)$, $h_3(x,y) = -f'(x)\cos(y_i)$ and application of Itô formula, as the Meyer process is $M(t)$ is equal to 0. See also Etheridge and Kurtz (2018), Appendix A.3, where a similar argument is discussed for the systems of SDEs with jumps. □

Proof[of Theorem 2.85] We begin by observing that $D(\widehat{A})$ is separable and closed under multiplication. The next step is to construct the transition function α in the Markov Mapping Theorem. Let $\mathcal{P}(\Omega)$ be the collection of probability measures on Ω. Let $\eta_y \in \mathcal{P}([0,2\pi)^\infty)$ be the product of independent uniform distributions on $[0,2\pi)$. We define

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

We observe that

$$\alpha L_x f(x,y) = L_x \alpha f(x,y). \quad (2.31)$$

Since η_y has been chosen in a way which guarantees that it is a stationary distribution for Y, we have, by definition,

$$\alpha L_y f(x,y) = 0. \quad (2.32)$$

Since D_0 is a collection of periodic functions, integration by parts shows that

$$\alpha L_{xy} f(x,y) = 0. \quad (2.33)$$

Combining (2.31), (2.32), (2.33),

$$\alpha\widehat{A}f(x,y) = A\alpha f(x,y).$$

Applying the Markov Mapping Theorem with $C = A$ and $B = \widehat{A}$, to every solution M to the martingale problem (2.25) there corresponds a solution $Z = \binom{M}{Y}$ to the martingale problem for $\hat{A}$ such that M has the same distribution as $\gamma \circ Z$. Since Z corresponds to a weak solution to (2.26), the result follows. □

2.7 Poisson random measures

In this section we would like to collect some facts about Poisson random measures. These objects, and some of their properties, play a crucial role for the results in Chapter 4. Most of the facts presented in this section have been recurrently stated in the papers on lookdown constructions, see, for example Kurtz and Rodrigues (2011), Etheridge and Kurtz (2018). For an excellent introductory presentation of the topic, which contains most of the proofs of the results stated in this section, we refer to Çınlar (2011), Chapter VI.

The following two Lemmas provide us with a collection of identities for Poisson random measures (see Definition 2.22).

Lemma 2.90. *Let ξ be a Poisson random measure with mean measure ν. Let $f \in L^1(\mathbb{R}^d, \nu)$. Then*

$$\mathbb{E}\Big[\exp\Big(\int f(z)\xi(\mathrm{d}z)\Big)\Big] = \exp\Big(\int (e^{f(x)} - 1)\nu(\mathrm{d}x)\Big)$$

Similarly, the expected value and variance of the integral with respect to a Poisson random measure is given by

$$\mathbb{E}\Big[\int f(z)\xi(\mathrm{d}z)\Big] = \int f(x)\nu(\mathrm{d}x) \quad Var\Big[\int f(z)\xi(\mathrm{d}z)\Big] = \int f^2(x)\nu(\mathrm{d}x)$$

Definition 2.91 (Conditionally Poisson system)**.** *Consider a counting measure ξ on $\mathbb{R}^d$. Let Ξ be a locally finite random measure on $\mathbb{R}^d$. We say that ξ is conditionally Poisson with Cox measure Ξ if, conditioned on Ξ, ξ is a Poisson random measure with mean measure Ξ.*

Conditionally Poisson systems are sometimes referred to as Cox processes. We notice that to check that a Poisson random measure ξ is actually a Cox process, it is enough to check that

$$\mathbb{E}\Big[\exp\Big(-\int_{\mathbb{R}^d} f\mathrm{d}\xi\Big)\Big] = \mathbb{E}\Big[\exp\Big(-\int_{\mathbb{R}^d}(1-e^f)\mathrm{d}\Xi\Big)\Big]$$

for all positive Borel-measurable functions f. Our main application of the presented theory is to show convergence of the particles systems to their high intensity limits. We need some more definitions, as the convergence of sequences of conditionally Poisson systems requires a rather exotic topology.

Consider a family of continuous functions $h_k : \mathbb{R}^d \to [0,1]$ such that

$$\bigcup_k S_{h_k} = \mathbb{R}^d,$$

where S_f denotes the support of f. Let $\mathcal{M}_{h_k}(\mathbb{R}^d)$ be the collection of Borel measures on $\mathbb{R}^d$ such that

$$\int_{\mathbb{R}^d} h_k \mathrm{d}\nu < \infty.$$

Let $\mathrm{d}\nu^k = h_k\mathrm{d}\nu$. The space $\mathcal{M}_{h_k}(\mathbb{R}^d)$ endowed with the topology of weak convergence of $\mathrm{d}\nu^k$ is metrizable. We observe that checking convergence in $\mathcal{M}_{h_k}(\mathbb{R}^d)$ is equivalent to checking convergence of $\int_{\mathbb{R}^d} f\mathrm{d}\nu_n$ for all bounded and continuous functions which satisfy

$$\int_{\mathbb{R}^d} f\mathrm{d}\nu^k < \infty \text{ for } f \le ch_k$$

for some constant $c > 0$. The space $\mathcal{M}_{h_k}(\mathbb{R}^d \times [0,\infty))$ can be defined in a similar way.

Theorem 2.92 (Kurtz and Rodrigues (2011), Theorem A.9)**.** *Let ξ_n be a sequence of conditionally Poisson random measures on $\mathbb{R}^d \times [0,\infty)$ with Cox measures $\{\Xi_n \times \Lambda\}$. Then $\xi_n \Rightarrow \xi$ in $\mathcal{M}_{h_k}(\mathbb{R}^d \times [0,\infty))$ if and only if $\Xi_n \Rightarrow \Xi$ in $\mathcal{M}_{h_k}(\mathbb{R}^d)$. If the limit exists, ξ is a conditionally Poisson random measure with Cox measure $\Xi \times \Lambda$.*

Chapter 3

Spatial Lambda-Fleming-Viot Model in a fluctuating environment

In this chapter we are concerned with one of the main results of the thesis, that is identification of the diffusion approximation for the Spatial Lambda-Fleming-Viot model in a fluctuating environment.

We begin by investigating the model in the absence of spatial structure in Section 3.1. The main aim of this investigation is an introduction of the main ideas of the proof, which are harder to follow in the spatial case. Section 3.2 describes the model, scaling and main results. Section 3.3 is devoted to the rather limited cases in which a duality argument is available. Section 3.4 contains a description of the 'tracer dynamics' (dynamics of a marked subpopulation). Section 3.5 is devoted to the proof of main result.

3.1 The non-spatial case

3.1.1 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), 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.

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 $\mathrm{d}t\otimes\nu(\mathrm{d}u)\otimes\sigma(\mathrm{d}\mathbf{s})$, 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*

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

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.

We should also mention that the results in a spirit similar to ours have been obtained for the Wright-Fisher model in a random environment in the recent paper Hutzenthaler et al. (2018)

3.1.2 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 $\mathrm{d}t \otimes \delta_{\bar{u}} \otimes \delta_{\mathbf{s}}$ 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

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

$$+\left[\frac{1-p}{1+\mathbf{s}_n p}\right] f((1-u_n)p,\zeta) - f(p,\zeta)\Big)$$
$$+\mathbf{1}_{\{\zeta=1\}} n\Big(\left[\frac{p}{1+\mathbf{s}_n(1-p)}\right] f((1-u_n)p+u_n,\zeta)$$
$$+\left[\frac{(1+\mathbf{s}_n)(1-p)}{1+\mathbf{s}_n(1-p)}\right] f((1-u_n)p,\zeta) - f(p,\zeta)\Big)$$
$$+ n^{\gamma}(\mathbb{E}_{\pi}[f(p,\cdot)] - f(p,\zeta)).$$

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

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

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

$$u_n = n^{-\frac{1}{2}}\bar{u} \quad \mathbf{s}_n = n^{-\frac{1}{2}+\alpha}\mathbf{s}, \qquad (3.2)$$

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

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

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 citekurtz:1973. We apply the generator (3.3) to test functions of the form

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

For this choice, we obtain

$$\begin{aligned}\mathcal{L}g(p,\zeta) = \mathcal{L}^{neu}f(p) + n^{\alpha}\Big(\mathcal{L}^{fsel}f\Big)(p,\zeta) + n^{-\delta}\mathcal{L}^{neu}\Big(\mathcal{L}_n^{fsel}f\Big)(p,\zeta)\\ + n^{\alpha-\delta}\mathcal{L}^{fsel}\Big(\mathcal{L}^{fsel}f\Big)(p,\zeta) - n^{\gamma-\delta}\Big(\mathcal{L}^{fsel}f\Big)(p,\zeta) + \mathcal{O}\Big(n^{-\frac{1}{2}+2\alpha}\Big),\end{aligned} \tag{3.5}$$

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

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

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

To evaluate the right hand side,

$$\begin{aligned}\mathcal{L}^{fsel}\Big(\mathcal{L}^{fsel}f\Big)(p,\zeta) = \mathcal{L}^{fsel}(-\zeta\bar{u}\mathbf{s}p(1-p)f'(p)) = \zeta^2\bar{u}^2\mathbf{s}^2p(1-p)\frac{d}{dp}\Big(p(1-p)\frac{d}{dp}f(p)\Big)\\ = \zeta^2\bar{u}^2\mathbf{s}^2p(1-p)(1-2p)f'(p) + \zeta^2\bar{u}^2\mathbf{s}^2p^2(1-p)^2f''(p).\end{aligned}$$

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 Miller (2012) 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 Kurtz (1992), which we recall later as Theorem 3.16. 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 càdalàg 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

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

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

3.2 Definition and scaling of the Spatial Lambda-Fleming-Viot process with fluctuating selection

In this section we first extend the Lambda-Fleming-Viot model with fluctuating selection of Section 3.1 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.

3.2.1 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 et al. (2014), which in turn borrows results from Véber and Wakolbinger (2015), 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 $\mathbb{R}^d$. This space of measures, which we denote by $\mathcal{M}_\lambda$, is equipped with the topology of vague convergence, under which it is compact. At any fixed time there is a density $w(t,\cdot) : \mathbb{R}^d \to [0,1]$ such that

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

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. 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 3.1 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 $p+\mathbf{s}p(1-p) = p(1-\mathbf{s})+\mathbf{s}(1-(1-p)^2)$ 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 $n \to \infty$.

Definition 3.5 (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*

$$\int_{(0,\infty)} r^d \int_{[0,1]} u\, \nu_r(\mathrm{d}u)\mu(\mathrm{d}r) < \infty. \tag{3.8}$$

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* $q(x,y)$ *be an element of* $C_b\Big(\mathbb{R}^d \times \mathbb{R}^d\Big)$ *and let* $\{\xi^{(m)}(\cdot)\}_{m\geq 0}$ *be a family of identically distributed random fields such that*

$$\mathbb{P}\Big[\xi^{(m)}(x) = -1\Big] = \frac{1}{2} = \mathbb{P}\Big[\xi^{(m)}(x) = +1\Big],$$
$$\mathbb{E}\Big[\xi^{(m)}(x)\xi^{(m)}(y)\Big] = q(x,y).$$

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* M_t *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 distribution*

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

(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 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*, independently, according to*

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

(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* $y \in B(x, r)$*,*

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

We have tacitly assumed that $w(t_-, l_i)$ is defined. It is, with probability one, so we declare that if it is not defined, then we resample and try again. For a construction of a

random field of Definition 3.5 we refer to Ma (2009), especially their Example 1. The arguments presented there require only a minor adaptation.

Existence of the SLFVFS is guaranteed by the methods of 2018, Etheridge et al. (2014). Indeed, we could have taken different measures μ and ν_r according to whether events are selective or neutral. It was convenient also to take the strength of selection to be constant in space and have its sign 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 3.8 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 'marked' according to the state of the environment at the time at which they occur.

In analogy with 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

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

The distribution of these Poisson point processes is invariant under the time reversal.

We suppose that at time 0 (which we think of as 'the present'), we sample k individuals from locations $x_1, \ldots, x_k$ and 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. It is convenient to consider this process in the space of finite point measures on $\mathbb{R}^d$, which we denote by $\mathcal{M}_p(\mathbb{R}^d)$.

Definition 3.6 (Ancestral selection graph)**.** *We first define a Markov process $(\Xi_t)_{t\geq 0}$ with càdlàg paths, taking values in $\mathcal{M}_p(\mathbb{R}^d)$, with the dynamics defined as follows. First, let*

$$\Xi_0 = \sum_i \delta_{\chi(s^i)}$$

be a finite point measure. We update Ξ as follows:

1. *At each $\tau \in \overleftarrow{\Pi}^{env}$, the environment is resampled;*
2. *(a) At each event $(t, x, r, u) \in \overleftarrow{\Pi}^{neu}$,*
    - *i. for each $\chi_{t-}^i \in B(x, r)$, independently mark the corresponding potential ancestor with probability u;*
    - *ii. if at least one lineage is marked, all marked lineages disappear (that is, their point masses are removed from Ξ_t) and are replaced by a single potential ancestor, whose location is drawn uniformly at random from within $B(x, r)$.*

    (b) At each event $(t, x, r, u) \in \overleftarrow{\Pi}^{sel}$:
    - *i. For each $\chi_{t-}^i \in B(x, r)$, independently mark the corresponding potential ancestor with probability u;*
    - *ii. 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 from locations $\chi_0^1, \ldots, \chi_0^j$, knowing the distribution of $w(T, x)$ at time T, 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 $\mathbb{P}[\chi_T^i = a] = w(T, \chi_T^i)$. 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 $\zeta = 1$ then all coalescing lineages are type a if and only if both parents are type a, otherwise they are type A; if $\zeta = -1$, 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), Neuhauser and Krone (1997), 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 3.7. *More formally, since the density of the SLFVFS is only defined Lebesgue-almost everywhere, the process of assigning types to the potential ancestral lineages at time t is only defined for Lebesgue almost every choice of $\chi_t^1, \ldots, \chi_t^{N_t}$ and, just as in Etheridge et al. (2014) Section 1.2, this duality must be defined 'weakly', that is by integrating against a test function. Since we do not use the duality in what follows, we refer the reader to Etheridge et al. (2014) for details.*

3.2.2 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 et al. (2014) 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 corresponding SPDE 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 3.1, 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 3.1). In order to be able to ignore terms of order $n\mathbf{s}_n^2 u_n$, we must now take $\alpha \in (0, 1/6)$.

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

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

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

Remark 3.8 (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* n*th 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* n*th stage, to ensure that the* $|\mathbf{s}_n(x)\xi_n(x)| < 1$*, 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

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

plus some regularity to reflect equation (3.10) 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 3.9. *From now on, fix $R \in (0, \infty)$ and $\bar{u} \in (0, 1)$ and take*

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

Just as in Etheridge et al. (2014), 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

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

and define the sequence of rescaled processes

$$w_n(t, x) = w(nt, n^{\frac{1}{3}}x), \quad \overline{w}_n(t, x) = \fint_{B_n(x)} w_n(t, y)\mathrm{d}y, \tag{3.9}$$

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

Theorem 3.10. *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 $q(x, y)$ that determines the environment satisfies*

$$|q(x, x) - q(x, y)| \leq C|x - y| \qquad \textit{for all } x, y \in \mathbb{R}^d. \tag{3.10}$$

Then the sequence $(\overline{M}^n_\cdot)_{n\geq 1}$ *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* $(M^\infty_t)_{t\geq 0}$*, writing* w^∞ *for a representative of the density of* M^∞*,*

1. *for dimension* $d=1$*,* w^∞ *is the process for which, for every* $F\in C^\infty_c(\mathbb{R})$ *and for every* $f\in C^2_c(\mathbb{R})$*,*

$$\begin{aligned}
&F(\langle w^\infty(t),f\rangle)-F(\langle w^\infty(0),f\rangle)\\
&-\int_0^t F'(\langle w^\infty(s),f\rangle)\Big\{\Big\langle w^\infty(s),\frac{\bar{u}\Gamma_R}{2}\Delta f\Big\rangle\\
&\qquad\qquad+V_R^2\bar{u}^2\mathbf{s}^2\langle w^\infty(s)(1-w^\infty(s))(1-2w^\infty(s)),f\rangle\Big\}\mathrm{d}s\\
&\quad-\int_0^t F''(\langle w^\infty(s),f\rangle)\Big\{\\
&\qquad\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Big[V_R^2\bar{u}^2\mathbf{s}^2q(x,y)w^\infty(s,x)(1-w^\infty(s,x))\\
&\qquad\qquad\times w^\infty(s,y)(1-w^\infty(s,y))f(x)f(y)\Big]\mathrm{d}x\mathrm{d}y\\
&\qquad\qquad+\frac{\bar{u}^2V_R^2}{2}\langle w^\infty(s)(1-w^\infty(s)),f^2\rangle\Big\}\mathrm{d}s \qquad (3.11)
\end{aligned}$$

   is a martingale;

2. *for dimension* $d\geq 2$*, for every* $F\in C^\infty_c(\mathbb{R})$ *and for every* $f\in C^2_c\left(\mathbb{R}^d\right)$*,*

$$\begin{aligned}
&F(\langle w^\infty(t),f\rangle)-F(\langle w^\infty(0),f\rangle)\\
&-\int_0^t F'(\langle w^\infty(s),f\rangle)\Big\{\Big\langle w^\infty(s),\frac{\bar{u}\Gamma_R}{2}\Delta f\Big\rangle\\
&\qquad\qquad+V_R^2\bar{u}^2\mathbf{s}^2\langle w^\infty(s)(1-w^\infty(s))(1-2w^\infty(s)),f\rangle\Big\}\mathrm{d}s\\
&\quad-\int_0^t F''(\langle w^\infty(s),f\rangle)\Big\{\\
&\qquad\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\Big[V_R^2\bar{u}^2\mathbf{s}^2q(x,y)w^\infty(s,x)(1-w^\infty(s,x))\\
&\qquad\qquad\times w^\infty(s,y)(1-w^\infty(s,y))f(x)f(y)\Big]\mathrm{d}x\mathrm{d}y\Big\}\mathrm{d}s \qquad (3.12)
\end{aligned}$$

   is a 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* (3.37)*.*

The proof of uniqueness in $d \geq 2$ uses a pathwise uniqueness result of Rippl and Sturm (2013) for the corresponding SPDE we apply Theorem 2.85 to show that any solution to the martingale problem (3.12) is actually a weak solution to the corresponding SPDE:

$$\mathrm{d}w^\infty = \left(\frac{\bar{u}\Gamma_R}{2}\Delta w^\infty + \bar{u}^2 V_R^2 \mathbf{s}^2 w^\infty(1-w^\infty)(1-2w^\infty)\right)\mathrm{d}t + \sqrt{2}\bar{u}V_R\mathbf{s}w^\infty(1-w^\infty)W(\mathrm{d}t,\mathrm{d}x), \quad (3.13)$$

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

$$\langle W(\phi)\rangle_t = t\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} q(x,y)\phi(x)\phi(y)\mathrm{d}x\mathrm{d}y. \quad (3.14)$$

The corresponding equation for dimension $d = 1$ is

$$dw^\infty = \left(\frac{\bar{u}\Gamma_R}{2}\Delta w^\infty + 4R^2\bar{u}^2\mathbf{s}^2 w^\infty(1-w^\infty)(1-2w^\infty)\right)dt + \sqrt{2}2R\bar{u}\mathbf{s}w^\infty(1-w^\infty)W(\mathrm{d}t,\mathrm{d}x) + 2R\bar{u}\sqrt{w^\infty(1-w^\infty)}\mathcal{W}(\mathrm{d}t,\mathrm{d}x), \quad (3.15)$$

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 (3.11) will be a weak solution to this SPDE, but we only have a proof of uniqueness of (3.15) in the special case in which W is also space-time white noise (in which case we can invoke the duality of Section 3.3).

3.3 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 SPDE being white in space as well as time. In order to obtain these duals, we transform the system in a way inspired by Blath et al. (2007).

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

$$\mathrm{d}X_t = \frac{1}{2}\bar{u}^2\mathbf{s}^2\Big(X_t^3 - X_t\Big)\mathrm{d}t + \bar{u}\sqrt{1 - X_t^2}\mathrm{d}B_t^1 + \frac{\sqrt{2}}{2}\bar{u}\mathbf{s}(1 - X_t^2)\mathrm{d}B_t^2, \tag{3.16}$$

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

Lemma 3.11. *The solution to the transformed equation* (3.16) *is dual to a branching annihilating process* $\{N_t\}_{t\geq 0}$ *with transitions*

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

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

2. $N \mapsto N - 2$ *at rate*

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

The duality relationship takes the form

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

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 the dual process will die out in finite time, (as it is essentially 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 3.12. *Suppose that $d = 1$ and w^∞ solves*

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

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 $\chi_1, \ldots, \chi_{N_t}$, and whose dynamics are described as follows:

1. *Each particle, independently, follows a Brownian motion in $\mathbb{R}$, with diffusion constant $\bar{u}\Gamma_R/2$;*
2. *Each particle, independently, splits into three at rate $\bar{u}\mathbf{s}^2$;*
3. *Each pair of particles annihilates, at a rate $\bar{u}^2(1+\mathbf{s}^2)$, measured by their intersection local time;*
4. *Each pair of particles replicates (i.e. is replaced by two identical pairs) at rate $\bar{u}^2\mathbf{s}^2$), 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 $w(0,\cdot) : \mathbb{R} \to [0,1]$, through

$$\mathbb{E}\left[\prod_{i=1}^{N_0} X_t(\chi_i(0)\right] = \mathbb{E}\left[\prod_{i=1}^{N_t} X_0(\chi_i(t)\right], \tag{3.17}$$

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

Tribe (1995) 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), see also Liang (2009). We also note that Birkner (2003)

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 3.13. *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. Passing with the size of the lattice to* 0 *leads directly to the equation of Lemma 3.12, see Doering et al. (2003) for a version with a single white noise, which adapts to our case without any significant changes. Moreover, we can apply the analogous transformation to the SPDE 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*

$$q(x,y)X_t(x)X_t(y)\Big(1-X_t(x)^2\Big)\Big(1-X_t(y)^2\Big).$$

We can rearrange the factors involving X as

$$\Big(X_t(x)^2X_t(y)^2-X_t(x)X_t(y)\Big)+(1-X_t(x)X_t(y)) \\ -\Big(X_t(x)^2-X_t(x)X_t(y)\Big)-\Big(X_t(y)^2-X_t(x)X_t(y)\Big). \quad (3.18)$$

If $q(x,y)>0$, then we can interpret the first two terms in (3.18) as branching and annihilating terms in a putative dual; if $q(x,y)<0$, 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) 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 (3.17), which it turns out is infinite.

3.4 Tracer dynamics

On passing to a SPDE 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), or, for a more mathematical approach, see Durrett and Fan (2016). 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 $v(t,x) \leq w(t,x)$.

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 3.14 (The SLFVFS with tracers)**.** *Let ζ, Π^{neu}, Π^{fsel}, Π^{env} be exactly as in Definition 3.5. 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 $\kappa \in \{0,1\}$ according to distribution*

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

(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

$$\mathbb{P}[\kappa_i = a^*] = v(t_-,l_i), \quad \mathbb{P}\Big[\kappa_i = a^\dagger\Big] = w(t_-,l_0) - v(t_-,l_0),$$
$$\mathbb{P}[\kappa_i = A] = 1 - w(t_-,l_i).$$

(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

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

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

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

Theorem 3.15. *Applying our previous scalings, let the population evolve under the assumptions of Theorem 3.10.*

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 $v^\infty(0,x) = \lim_{n\to\infty} \overline{v}_n(0,x)$. 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 SPDEs which in* $d \geq 2$ *takes the form*

$$\begin{aligned}
\mathrm{d}w^\infty &= \Big(\frac{\bar{u}\Gamma_R}{2}\Delta w^\infty + V_R^2\bar{u}^2\mathbf{s}^2 w^\infty(1-w^\infty)(1-2w^\infty)\Big)\mathrm{d}t + \sqrt{2}V_R\bar{u}\mathbf{s}w^\infty(1-w^\infty)W(\mathrm{d}t,\mathrm{d}x),\\
dv^\infty &= \Big(\frac{\bar{u}\Gamma_R}{2}\Delta v^\infty + V_R^2\bar{u}^2\mathbf{s}^2 v^\infty(1-w^\infty)(1-2w^\infty)\Big)\mathrm{d}t + \sqrt{2}V_R\bar{u}\mathbf{s}v^\infty(1-w^\infty)W(\mathrm{d}t,\mathrm{d}x),
\end{aligned}$$

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*

$$\begin{aligned}
\mathrm{d}w^\infty =& \Big(\frac{\bar{u}\Gamma_R}{2}\Delta w^\infty + 4R^2\bar{u}^2\mathbf{s}^2 w^\infty(1-w^\infty)(1-2w^\infty)\Big)\mathrm{d}t + \sqrt{2}2R\bar{u}w^\infty(1-w^\infty)W(\mathrm{d}t,\mathrm{d}x)\\
&+ 2R\bar{u}\sqrt{v^\infty(1-w^\infty)}\mathcal{W}^0(\mathrm{d}t,\mathrm{d}x) + 2R\bar{u}\sqrt{(w-v)^\infty(1-w^\infty)}\mathcal{W}^1(\mathrm{d}t,\mathrm{d}x),\\
\mathrm{d}v^\infty =& \Big(\frac{\bar{u}\Gamma_R}{2}\Delta v^\infty + 4R^2\bar{u}^2\mathbf{s}^2 v^\infty(1-w^\infty)(1-2w^\infty)\Big)\mathrm{d}t + \sqrt{2}2R\bar{u}\mathbf{s}v^\infty(1-w^\infty)W(\mathrm{d}t,\mathrm{d}x)\\
&+ 2R\bar{u}\sqrt{v^\infty(1-w^\infty)}\mathcal{W}^0(\mathrm{d}t,\mathrm{d}x) + 2R\bar{u}\sqrt{v^\infty(w-v^\infty)}\mathcal{W}^2(\mathrm{d}t,\mathrm{d}x),
\end{aligned}$$

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 3.10, 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 SPDE'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). In the case of one dimension and genic selection, Durrett and Fan (2016) obtain a pair of SPDE's of the form

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

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 3.1, 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

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

From this we see that the SPDEs in Theorem 3.15 are of precisely the form that we should expect.

3.5 Proof of Theorem 3.10

The proof of Theorem 3.10 will rest on Theorem 2.1 of Kurtz (1992) (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 $\mu([0,t)\times E)=t$.

Theorem 3.16 (Kurtz (1992))**.** *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 $K\subset E$ such that*

$$\inf_n \mathbb{P}[X_n(t)\in K, t\le T]\ge 1-\epsilon, \tag{3.19}$$

and assume that $\{Y_n(t): t\ge 0, n=1,2,\ldots\}$ 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

$$f(X_n(t))-\int_0^t Af(X_n(s),Y_n(s))\mathrm{d}s+\epsilon_n^f(t) \tag{3.20}$$

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

$$\sup_n \mathbb{E}\left[\int_0^T |Af(X_n(t), Y_n(t)|^p \mathrm{d}t\right] < \infty \tag{3.21}$$

and

$$\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\leq T} |\epsilon_n^f(t)|\right] = 0. \tag{3.22}$$

Let Γ_n be the $\mathcal{M}(E_2)$-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 filration $\{\mathcal{G}_t\}$ such that

$$f(X(t)) - \int_0^t \int_{E_2} Af(X(s), y)\Gamma(\mathrm{d}s\times \mathrm{d}y) \tag{3.23}$$

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

As a particular case, Kurtz (1992) provides the following example.

Example 3.17 (Kurtz (1992))**.** *Suppose Y is stationary and ergodic, and Y_n in Theorem 3.16 is given by $Y_n(t) \equiv Y(nt)$. Then $\Gamma = \Lambda \times \pi$ where Λ denotes Lebesgue measure and π is the marginal distribution for Y. Consequently, under the assumptions of Theorem 3.16, 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 that of Theorem 3.16. The sequence of measure-valued evolutions $(\overline{M}_t^n)_{t\geq 0}$ corresponds to the process X_n, while the sequence of environments corresponds to the process Y_n. Notice that Y_n defined that way satisfies the assumptions of Example 3.17, and therefore those of Theorem 3.16. The domain of the generator A is given by the closure of

$$\mathcal{D}(A) = \left\{\psi_{F,f} = F(\langle\cdot, w\rangle) : f \in C_c^\infty(\mathbb{R}^d), F \in C^1(\mathbb{R})\right\}. \tag{3.24}$$

Since the state space of $\left(\overline{M}_t^n\right)_{t\geq 0}$ is compact, the compact containment condition (3.19) is satisfied.

The key step towards the application of Theorem 3.16 is to identify the generator A, which leads to decomposition (3.20). This consists of two steps, approximation of the parts of the SLFVFS generator corresponding to neutral events and the second, which deals with terms corresponding to fluctuating selection. These procedure is the subject of Subsection 3.5.1.

We begin by transforming the generator of SLFVFS into a form which allows to consider the process of averaged densities $\overline{w}_t^n$. We then use approach of Kurtz (1973), namely the 'separation of timescales' trick, which allows us to split the generator into parts corresponding to neutral and selective events in the correct way. This formal limit evaluation is achieved in (3.34). This step is analogous to the strategy of Section 3.1. We then proceed with the analysis of the two parts of the generator. The first one, which corresponds to the neutral events, will closely follow Etheridge et al. (2014); the novel step is the second, which deals with the terms corresponding to fluctuating selection. Our analysis not only identify the correct form of A and ϵ^f (see (3.46) and (3.47) respectively), but also provides a uniform bound on A, which shows that (3.21) is satisfied.

Having identified A, and ϵ^f, we show that (3.22) holds. This bound relies on the Lipschitz bound (3.10) on the correlation kernel $g(x, y)$. We finish Subsection 3.5.1 with discussion of the necessity of the estimates on the continuity of the densities of the limiting process, which we state as Proposition 3.20, which we prove in Subsection 3.5.3. This leads to a complete characterisation of limits of the sequence of averaged processes $\overline{w}_n$ as solutions to a martingale problem. We discuss the uniqueness of solutions to that martingale problem in Subsection 3.5.2

3.5.1 Identifying the limit

In what follows C_f is a constant that depends only on f, but may vary from line to line.

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 et al. (2014) 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

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

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

$$\begin{aligned}\phi_f(x) &= f(x) + \frac{n^{\frac{d}{3}}}{V_R}\int_{B_n(x)}(y-x)\nabla f(x)\mathrm{d}y + \mathcal{O}\left(\frac{n^{\frac{d}{3}}}{V_R}\int_{B_n(x)\cap S_f}|x-y|^2\mathrm{d}x\right)\\ &= f(x) + \mathcal{O}\left(n^{-\frac{2}{3}}\right)\mathbf{1}_{B_n(x)\cap S_f\neq\emptyset},\end{aligned} \tag{3.25}$$

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

$$\langle \overline{w}_n, f\rangle = \langle w_n, \phi_f\rangle = \langle w_n, f\rangle + \mathcal{O}(n^{-\frac{2}{3}}). \tag{3.26}$$

It will be convenient to have some notation.

Notation 3.18. *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*

$$\begin{aligned}\theta^+_{x,R,\bar{u}}(w) &= \mathbf{1}_{\{\mathbb{R}^d\setminus B(x,R)\}}w + \mathbf{1}_{\{B(x,R)\}}((1-\bar{u})w + \bar{u}),\\ \theta^-_{x,R,\bar{u}}(w) &= \mathbf{1}_{\{\mathbb{R}^d\setminus B(x,R)\}}w + \mathbf{1}_{\{B(x,R)\}}(1-\bar{u})w.\end{aligned} \tag{3.27}$$

Using this notation,

$$\langle \theta^{+}_{x,R_n,u_n}(w), \phi_f \rangle - \langle w, \phi_f \rangle = u_n \Big\langle \mathbf{1}_{B_n(x)}(1-w), \phi_f \Big\rangle, \tag{3.28}$$

$$\langle \theta^{-}_{x,R_n,u_n}(w), \phi_f \rangle - \langle w, \phi_f \rangle = -u_n \Big\langle \mathbf{1}_{B_n(x)} w, \phi_f \Big\rangle. \tag{3.29}$$

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

$$\psi_{F,f}(w_n) = F(\langle \overline{w}_n, f \rangle) = F(\langle w_n, \phi_f \rangle), \tag{3.30}$$

where $F \in C^{\infty}(\mathbb{R})$, $f \in C^{\infty}_c(\mathbb{R}^d)$. Notice that these are independent of the environment ζ. Recall that the analogous functions were our starting point in Section 3.1. Since the support S_f of f is compact, it is only overlapped by events at a finite rate and so, writing $\mathcal{L}F(\langle w, \phi_f \rangle)$, 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

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

We recognise the generator of the neutral spatial Lambda-Fleming-Viot process 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 et al. (2014) 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

$$\begin{aligned}\mathcal{L}_n F(\langle w_n, \phi_f\rangle) = n^{1+\frac{d}{3}}(1-\mathbf{s}_n)\Big(&\int_{\mathbb{R}^d} \overline{w}_n(x)\Big[F(\langle \theta^+_{x,R_n,u_n}(w_n), \phi_f\rangle) - F(\langle w_n, \phi_f\rangle)\Big]\mathrm{d}x \\ &+ \int_{\mathbb{R}^d}\{1-\overline{w}_n(x)\}\Big[F(\langle \theta^-_{x,R_n,u_n}(w_n), \phi_f\rangle) - F(\langle w_n, \phi_f\rangle)\Big]\mathrm{d}x\Big) \\ + \mathbf{s}_n \int_{\mathbb{R}^d}&\Big\{\overline{w}_n^2(x) + (1-\zeta(x))\overline{w}_n(x)(1-\overline{w}_n(x))\Big\} \\ &\times\Big[F(\langle \theta^+_{x,R_n,u_n}(w_n), \phi\rangle) - F(\langle w_n, \phi_f\rangle)\Big]\mathrm{d}x \\ + \mathbf{s}_n \int_{\mathbb{R}^d}&\Big\{(1-\overline{w}_n(x))^2 + (1+\zeta(x))\overline{w}_n(x)(1-\overline{w}(x))\Big\} \\ &\times\Big[F(\langle \theta^-_{x,R_n,u_n}(w_n), \phi_f\rangle) - F(\langle w_n, \phi_f\rangle)\Big]\mathrm{d}x.\end{aligned}$$

By analogy with (3.6), we write this as

$$\mathcal{L}_n F(\langle w_n, \phi_f\rangle) = (1-\mathbf{s}_n)\mathcal{L}_n^{neu} F(\langle w_n, \phi\rangle) + n^{\alpha}\mathcal{L}_n^{fsel} F(\langle w_n, \phi\rangle), \tag{3.31}$$

where

$$\begin{aligned}\mathcal{L}_n^{neu} F(\langle w_n, \phi_f\rangle) = n^{1+\frac{d}{3}}\Big(&\int_{\mathbb{R}^d} \overline{w}_n(x)\Big[F(\langle \theta^+_{x,R_n,u_n}(w_n), \phi_f\rangle) - F(\langle w_n, \phi_f\rangle)\Big]\mathrm{d}x \\ &+ \int_{\mathbb{R}^d}\{1-\overline{w}_n(x)\}\Big[F(\langle \theta^-_{x,R_n,u_n}(w_n), \phi_f\rangle) - F(\langle w_n, \phi_f\rangle)\Big]\mathrm{d}x\Big),\end{aligned} \tag{3.32}$$

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

$$\begin{aligned}\mathcal{L}_n^{fsel} F(\langle w_n, \phi_f\rangle) = n^{1+\frac{d}{3}}\tilde{\mathbf{s}}_n \int_{\mathbb{R}^d}&\{\overline{w}_n(x) - \zeta(x)\overline{w}_n(x)(1-\overline{w}_n(x))\} \\ &\times\Big[F(\langle \theta^+_{x,R_n,u_n}(w_n), \phi_f\rangle) - F(\langle w_n, \phi_f\rangle)\Big]\mathrm{d}x \\ + n^{1+\frac{d}{3}}\tilde{\mathbf{s}}_n \int_{\mathbb{R}^d}&\{(1-\overline{w}_n(x)) + \zeta(x)\overline{w}_n(x)(1-\overline{w}_n(x))\} \\ &\times\Big[F(\langle \theta^-_{x,R_n,u_n}(w_n), \phi_f\rangle) - F(\langle w_n, \phi_f\rangle)\Big]\mathrm{d}x.\end{aligned}$$

Just as in the non-spatial case, in order to reduce the generator to the form (3.20), we appeal to the approach of Kurtz (1973). Entirely analogously to our calculations of

Section 3.1, we evaluate the generator on test functions of the form

$$G(\langle w_n, \phi_f\rangle, \zeta) = F(\langle w_n, \phi_f\rangle) + n^{-\alpha}(\mathcal{L}_n^{fsel} F(\langle \cdot, \phi_f\rangle))(w_n, \zeta), \tag{3.33}$$

where $F \in C^\infty(\mathbb{R})$ and $f \in C_c^2(\mathbb{R}^d)$. The second term on the right hand side will form part of ϵ_n^f in (3.20). 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 $n^{2\alpha}$.

Calculating exactly as in the non-spatial case, we find

$$\begin{aligned}\mathcal{L}_n G(\langle w_n, \phi_f\rangle, \zeta) + \mathcal{O}(n^{-\alpha}) &= \mathcal{L}_n^{neu} F(\langle w_n, \phi_f\rangle) + \mathcal{L}_n^{fsel}\Big(\mathcal{L}_n^{fsel} F(\langle \cdot, \phi_f\rangle)(w_n, \zeta)\Big) \\ &:= \mathcal{L}_n^{neu} F(\langle w_n, \phi_f\rangle, \zeta) + \mathcal{G}_n(F, \langle w_n, \phi_f\rangle, \zeta). \end{aligned} \tag{3.34}$$

In this notation, for $F \in C^\infty(\mathbb{R})$ and $f \in C_c^\infty(\mathbb{R}^d)$

$$F(\langle w_n(t), \phi_f\rangle) - \int_0^t [\mathcal{L}_n^{neu} F(\langle w_n(s), \phi_f\rangle) + \mathcal{G}_n(F, \langle w_n(s), \phi_f\rangle)]\mathrm{d}s$$

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 et al. (2014) and so we recall the strategy, but omit many of the details.

Taking the expression (3.32) for $\mathcal{L}_n^{neu}$, we perform a Taylor expansion of F about $\langle w_n, \phi_f\rangle$ to obtain

$$\int_0^t \mathcal{L}_n^{neu} F(\langle w_n(s), \phi_f\rangle)\mathrm{d}s = (1 - \mathbf{s}_n)\int_0^t [A_n(s) + B_n(s) + C_n(s)]\mathrm{d}s,$$

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

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

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

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

where we have used (3.28) and (3.29) and C_n corresponds to the third order term in the Taylor expansion and the error coming from application of (3.26).

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

$$\begin{aligned}
A_n(s) =& u_n n^{1+\frac{d}{3}} F'(\langle \overline{w}_n(s), f\rangle) \int_{\mathbb{R}^d} \Big\{ \overline{w}_n(s,x) \langle \mathbf{1}_{B_n(x)}, \phi_f \rangle - \langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f \rangle \Big\} \mathrm{d}x \\
=& u_n n^{1+\frac{d}{3}} F'(\langle \overline{w}_n(t), f\rangle) \int_{\mathbb{R}^d} \Bigg[\frac{n^{\frac{d}{3}}}{V_R} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|<R_n\}} w_n(s,y) \mathbf{1}_{\{|x-z|<R_n\}} \phi_f(z) \mathrm{d}y \mathrm{d}z \\
& - \int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|<R_n\}} w_n(s,y) \phi_f(y) \mathrm{d}y \Bigg] \mathrm{d}x \\
=& u_n n^{1+\frac{d}{3}} F'(\langle \overline{w}_n(t), f\rangle) \int_{\mathbb{R}^d} \frac{n^{\frac{d}{3}}}{V_R} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|<R_n\}} w_n(s,y) \mathbf{1}_{\{|x-z|<R_n\}} \\
& (\phi_f(z) - \phi_f(y)) \mathrm{d}y \mathrm{d}z \mathrm{d}x.
\end{aligned}$$

Now Taylor-expand ϕ_f around y to obtain

$$\begin{aligned}
A_n(s) = u_n n^{1+\frac{d}{3}} F'(\langle \overline{w}_n(s), f\rangle) \int_{\mathbb{R}^d} \Bigg[\frac{n^{\frac{d}{3}}}{V_R} \int \int \mathbf{1}_{\{|x-y|<R_n\}} w_n(s,y) \mathbf{1}_{\{|x-z|<R_n\}} \\
\times \Big[D\phi_f(y)(z-y) + \frac{1}{2}(z-y)\mathbf{H}(y)\phi_f(z-y) + \mathcal{O}\Big(|z-y|^3\Big) \mathbf{1}_{y \in S_f} \Big] \Bigg] \mathrm{d}y \mathrm{d}z \mathrm{d}x, \quad (3.35)
\end{aligned}$$

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 (3.35) is equal to

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

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). The analogue of (3.25) with ϕ_f

replaced by $\frac{\partial^2\phi_f}{\partial y_i^2}$ is

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

and (using (3.26) with Δf in place of f) we deduce that

$$\begin{aligned}
A_n(s) &= u_n n^{1+\frac{d}{3}} F'(\langle\overline{w}_n(s), f\rangle) \int_{\mathbb{R}^d} \frac{n^{\frac{d}{3}}}{V_R} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_{\{|x-y|<R_n\}}\mathbf{1}_{\{|x-z|<R_n\}} w_n(s,y) \\
&\qquad \times \frac{1}{2}\sum_{i=1}^d (z_i-y_i)^2 \frac{\partial^2}{\partial y_i^2} f(y)\mathrm{d}y\mathrm{d}z\mathrm{d}x + \mathcal{O}\Big(n^{-\frac{2}{3}}C\Big) \\
&= \frac{\bar{u}\Gamma_R}{2} F'(\langle\overline{w}_n(s), f\rangle) \int_{\mathbb{R}^d} w_n(s,y)\Delta f \mathrm{d}y + \mathcal{O}\Big(n^{-\frac{2}{3}}C\Big) \\
&= \frac{\bar{u}\Gamma_R}{2} F'(\langle\overline{w}_n(s), f\rangle)\langle\overline{w}_n(s), \Delta f\rangle + \mathcal{O}\Big(n^{-\frac{2}{3}}C\Big)
\end{aligned} \tag{3.36}$$

where

$$\Gamma_R = \frac{1}{V_R}\int_{B(0,R)}\int_{B(x,R)} (z_1)^2 \mathrm{d}z\mathrm{d}x, \tag{3.37}$$

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:

$$\begin{aligned}
B_n(s) &= u_n^2 n^{1+\frac{d}{3}} \frac{F''(\langle\overline{w}_n(s), f\rangle)}{2} \\
&\quad \times \int_{\mathbb{R}^d} \overline{w}_n(s)\Big\langle \mathbf{1}_{B_n(x)}(1-w_n(s)), \phi_f\Big\rangle^2 - (1-w_n(s))\Big\langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f\Big\rangle^2 \mathrm{d}x \\
&\qquad = u_n^2 n^{1+\frac{d}{3}} \frac{F''(\langle\overline{w}_n(s), f\rangle)}{2} \int_{\mathbb{R}^d} \{\overline{w}_n(s)\Big\langle \mathbf{1}_{B_n(x)}, \phi_f\Big\rangle^2 \\
&\qquad - 2\overline{w}_n(s)\Big\langle \mathbf{1}_{B_n(x)}, \phi_f\Big\rangle\Big\langle w_n(s)\mathbf{1}_{B_n(x)}, \phi_f\Big\rangle + \Big\langle w_n(s)\mathbf{1}_{B_n(x)}, \phi_f\Big\rangle^2\Big\}\mathrm{d}x \\
&= u_n^2 n^{1+\frac{d}{3}} \frac{F''(\langle\overline{w}_n(s), f\rangle)}{2} \int_{\mathbb{R}^d} \Big\{\overline{w}_n(s)\Big\langle \mathbf{1}_{B_n(x)}, \phi_f\Big\rangle\Big(\Big\langle \mathbf{1}_{B_n(x)}, \phi_f\Big\rangle - \Big\langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f\Big\rangle\Big) \\
&\qquad + \Big\langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f\Big\rangle\Big(\Big\langle \mathbf{1}_{B_n(x)}, \phi_f\Big\rangle - \overline{w}_n(s)\Big\langle \mathbf{1}_{B_n(x)} w_n(s), \phi_f\Big\rangle\Big)\Big\}\mathrm{d}x \\
&\qquad \leq u_n^2 n^{1+\frac{d}{3}} \frac{F''(\langle\overline{w}_n(s), f\rangle)}{2} \int_{S_f} 2\mathtt{Vol}(B_n(x))^2 \|f\|_\infty^2 \mathrm{d}x \leq C n^{\frac{1-d}{3}},
\end{aligned} \tag{3.38}$$

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 particular, in $d = 1$, if we specify $F(x) = x^2$, an algebraic computation and an application of (3.26) allows us to specify the quadratic variation to

$$\int_0^t \bar{u}^2 V_R^2 \langle w^\infty(s)(1 - w^\infty(s)), f^2 \rangle \mathrm{d}s, \tag{3.39}$$

see pp. 30-31 of Etheridge et al. (2014) for the details.

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 $\mathcal{L}_n^{fsel} F(\langle \cdot, \phi_f \rangle)(w_n, \zeta)$. It is here that we must diverge from Etheridge et al. (2014).

As before, although the test function $F(\langle w_n, \phi_f \rangle)$ 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 $\mathcal{L}_n^{fsel}$. We take a Taylor expansion of F about $\langle w_n, \phi_f \rangle$ to obtain

$$\begin{aligned}
&\mathcal{L}_n^{fsel} F(\langle w_n, \phi_f \rangle) + \mathcal{O}\Big(n^{-\frac{1}{3}} C_f\Big) \\
&= n^{1+\frac{d}{3}} \tilde{\mathbf{s}}_n u_n \Big[\int_{\mathbb{R}^d} \{\overline{w}_n(x) - \zeta(x)\overline{w}_n(x)(1 - \overline{w}_n(x))\} F'(\langle w_n, \phi_f \rangle) \langle \mathbf{1}_{B_n(x)}(1 - w_n), \phi_f \rangle \mathrm{d}x \\
&- \int_{\mathbb{R}^d} \{(1 - \overline{w}_n(x)) + \zeta(x)\overline{w}_n(x)(1 - \overline{w}_n(x))\} F'(\langle w_n, \phi_f \rangle) \langle \mathbf{1}_{B_n(x)} w_n, \phi_f \rangle \mathrm{d}x \Big],
\end{aligned} \tag{3.40}$$

(plus lower order terms) where we have used (3.28) and (3.29).

Now using our scaling, (3.25) and (3.26), we obtain

$$\begin{aligned}
&\mathcal{L}_n^{fsel} F(\langle \overline{w}_n, \phi_f \rangle) \\
&= n^{\frac{d}{3}} \mathbf{s} \bar{u} F'(\langle w_n, \phi_f \rangle) \Big[\int_{\mathbb{R}^d} \{\overline{w}_n(x) - \zeta(x)\overline{w}_n(x)(1 - \overline{w}_n(x))\} \\
&\qquad\qquad \times \Big\langle \mathbf{1}_{B_n(x)}(1 - w_n), f + \mathcal{O}(||x - y||) \Big\rangle \mathrm{d}x \\
&- \int_{\mathbb{R}^d} \{(1 - \overline{w}_n(x)) + \zeta(x)\overline{w}_n(x)(1 - \overline{w}_n(x))\} \Big\langle \mathbf{1}_{B_n(x)} w_n, [f + \mathcal{O}(|x - y|)] \Big\rangle \mathrm{d}x \Big] + \mathcal{O}\Big(n^{-\frac{1}{3}}\Big) \\
&= -\mathbf{s} \bar{u} V_R F'(\langle w_n, \phi_f \rangle) \int_{\mathbb{R}^d} \zeta(x)\overline{w}_n(x)(1 - \overline{w}_n(x)) f(x) \mathrm{d}x + \mathcal{O}\Big(n^{-\frac{1}{3}}\Big) \\
&= -\mathbf{s} \bar{u} V_R F'(\langle w_n, \phi_f \rangle) H_f(\overline{w}_n, \zeta(\cdot)) + \mathcal{O}\Big(n^{-\frac{1}{3}}\Big),
\end{aligned} \tag{3.41}$$

where

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

We now need to evaluate $\mathcal{L}_n^{fsel}$ on the product in (3.41).

$$\begin{aligned}
&\mathcal{L}_n^{fsel}\Big(\mathcal{L}_n^{fsel}F(\langle\cdot,\phi_f\rangle)(w_n,\zeta)\Big) + \mathcal{O}\Big(n^{-\frac{1}{3}}\Big) \\
&\quad = -\mathbf{s}\bar{u}V_R\mathcal{L}_n^{fsel}(F'(\langle w_n,\phi_f\rangle))H_f(\overline{w}_n,\zeta(\cdot)) - \mathbf{s}\bar{u}V_R F'(\langle w_n,\phi_f\rangle)\mathcal{L}_n^{fsel}(H_f(\overline{w}_n,\zeta(\cdot)))
\end{aligned} \tag{3.42}$$

As in the non-spatial case, these two terms will correspond to the diffusion and drift terms respectively in the SPDE.

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

$$\begin{aligned}
&\left(\overline{w}_n(x) + \frac{n^{d/3}}{V_R}\int_{B_n(y)} \mathbf{1}_{\{|z-x|<R_n\}}u_n(1-w_n(z))\mathrm{d}z\right)^2 \\
&\qquad = \overline{w}_n(x)^2 + 2u_n\overline{w}_n(x)\frac{n^{d/3}}{V_R}\int_{B_n(y)}\fint_{B_n(y)} \mathbf{1}_{\{|z-x|<R_n\}}(1-w_n(z))\mathrm{d}z \\
&\qquad\qquad + \mathcal{O}(u_n^2\mathbf{1}_{\{|x-y|<2R_n\}}).
\end{aligned} \tag{3.43}$$

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 (3.42), 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 (3.43). We shall also use continuity of $\overline{w}_n(x)$. As shown in the proof of

Proposition 3.20 (see (3.5.3), (3.51)), for any fixed constant $\lambda > 0$,

$$|\overline{w}_n(x) - \overline{w}_n(y)| \le Cn^{-1/12} e^{\lambda|x|}, \qquad \text{for } |x-y| < 2R_n.$$

We can now observe that the second term in (3.42) can be written as

$$\begin{aligned}
&\mathcal{L}_n^{fsel}\Big(\int_{\mathbb{R}^d} \zeta(x)\overline{w}_n(x)(1-\overline{w}_n(x))f(x)\mathrm{d}x\Big) \\
&\quad = n^{1+\frac{d}{3}}\tilde{\mathbf{s}}_n \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \{\overline{w}_n(y) - \zeta(y)\overline{w}_n(y)(1-\overline{w}_n(y))\}\zeta(x) \\
&\qquad\qquad \times \frac{n^{d/3}u_n}{V_R}\mathbf{1}_{\{|x-z|<R_n\}}\mathbf{1}_{\{|y-z|<R_n\}}(1-w_n(z))(1-2\overline{w}_n(x))f(x)\mathrm{d}z\mathrm{d}y\mathrm{d}x \\
&\quad - n^{1+\frac{d}{3}}\tilde{\mathbf{s}}_n \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \{(1-\overline{w}_n(y)) + \zeta(y)\overline{w}_n(y)(1-\overline{w}_n(y))\}\zeta(x) \\
&\qquad \times \frac{n^{d/3}u_n}{V_R}\mathbf{1}_{\{|x-z|<R_n\}}\mathbf{1}_{\{|y-z|<R_n\}}w_n(z)(1-2\overline{w}_n(x))f(x)\mathrm{d}z\mathrm{d}y\mathrm{d}x + \mathcal{O}(n^{-\frac{1}{3}})
\end{aligned} \tag{3.44}$$

We can further simplify this expression to

$$\begin{aligned}
&\mathcal{L}_n^{fsel}\Big(\int_{\mathbb{R}^d} \zeta(x)\overline{w}_n(x)(1-\overline{w}_n(x))f(x)\mathrm{d}x\Big) \\
&= -n^{1+\frac{d}{3}}\tilde{\mathbf{s}}_n \frac{n^{d/3}u_n}{V_R}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \zeta(x)\zeta(y)\mathbf{1}_{\{|x-z|<R_n\}}\mathbf{1}_{\{|y-z|<R_n\}} \\
&\qquad\qquad \times (1-2\overline{w}_n(x))f(x)\overline{w}_n(y)(1-\overline{w}_n(y))\mathrm{d}z\mathrm{d}y\mathrm{d}x \\
&+ \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \Big\{\zeta(x)\mathbf{1}_{\{|x-z|<R_n\}}\mathbf{1}_{\{|y-z|<R_n\}} \\
&\qquad\qquad \times (1-2\overline{w}_n(x))f(x)(\overline{w}_n(y)-\overline{w}_n(z))\Big\}\mathrm{d}z\mathrm{d}y\mathrm{d}x + \mathcal{O}(n^{-\frac{1}{3}}) \\
&= -\bar{u}\mathbf{s}V_R\int \zeta(x)^2\overline{w}_n(x)(1-\overline{w}_n(x))(1-2\overline{w}_n(x))f(x)\mathrm{d}x \\
&+ \bar{u}\mathbf{s}V_R\int \fint_{B_n(x)} \Big(\zeta(x)^2 - \zeta(x)\zeta(y)\Big)\overline{w}_n(x)(1-\overline{w}_n(x))(1-2\overline{w}_n(x))f(x)\mathrm{d}y\mathrm{d}x \\
&\qquad\qquad + \mathcal{O}(Cn^{-\frac{1}{12}}),
\end{aligned} \tag{3.45}$$

where we have used the continuity estimates on $\overline{w}_n$ to obtain the error in (3.45).

Since $F'(\langle w_n, \phi_f\rangle)$ is independent of ζ, the first term in (3.42) 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 3.16 to be

$$AF(\langle\overline{w},f\rangle) = F'(\langle\overline{w},f\rangle)\Big\{\langle\overline{w},\frac{\bar{u}\Gamma_R\Delta f}{2}\rangle + \langle\bar{u}^2\mathbf{s}^2V_R^2\overline{w}(1-\overline{w})(1-2\overline{w}),f\rangle\Big\}$$
$$+ F''(\langle\overline{w},f\rangle)\Big\{\bar{u}^2\mathbf{s}^2V_R^2\langle\zeta\overline{w}(1-\overline{w},f\rangle^2 + \mathbf{1}_{d=1}\frac{\bar{u}^2V_R^2}{2}\langle\overline{w}(1-\overline{w}),f^2\rangle\Big\}, \quad (3.46)$$

and

$$\epsilon_n^f(t) = \bar{u}^2\mathbf{s}^2V_R^2\int_0^t\Big\{F'(\langle\overline{w},f\rangle)\int_{\mathbb{R}^d}\fint_{B_n(x)}\Big(\zeta(x)^2-\zeta(x)\zeta(y)\Big)$$
$$\times\,\overline{w}_n(x)(1-\overline{w}_n(x))(1-2\overline{w}_n(x))f(x)\Big\}\mathrm{d}y\mathrm{d}x\mathrm{d}s + Ct\mathcal{O}(n^{-\frac{1}{12}}+n^{-\alpha}), \quad (3.47)$$

where the term of order $n^{-\alpha}$ comes from the second term on the right of (3.33). The additional term in A in $d=1$ stems from the term $B_n(s)$ in (3.38) and we have again used the continuity estimates on $\overline{w}$. Since A is uniformly bounded, (3.21) is obviously satisfied. To check that $\mathbb{E}[\sup_{t\le T}|\epsilon_n^f(t)|]\to 0$ as $n\to\infty$, we must control the integral term in (3.47).

Now for $|x-y|<R_n$, by (3.10),

$$\mathbb{E}[\zeta(x)^2-\zeta(x)\zeta(y)] = \Big(g(x,x)-g(x,y)\Big)\le CR_n$$

so that for any given realisation of the environment,

$$\mathbb{E}\Big[\int\fint_{B_n(x)}\Big(\zeta(x)^2-\zeta(x)\zeta(y)\Big)\overline{w}_n(x)(1-\overline{w}_n(x))(1-2\overline{w}_n(x))f(x)\mathrm{d}x\Big]\le Cn^{-1/3}.$$

On the other hand the integrand is uniformly bounded by a constant, so that writing S for the supremum of the modulus of the integral over $n^{2\alpha}$ realisations of the environment (recall that τ_i denote the times of the environmental switches),

$$\mathbb{E}[S] = \mathbb{E}\Big[\sup_{t<T}\Big\{\sum_{m=0}^{n^{2\alpha}-1}\int_{\tau_m\wedge t}^{\tau_{m+1}\wedge t}|\epsilon_n^f(s)|\mathrm{d}s\Big\}\Big]$$
$$\le n^{2\alpha}\mathbb{E}\Big[\sup_m\int_{\tau_m\wedge t}^{\tau_{m+1}\wedge t}|\epsilon_n^f(s)|\mathrm{d}s\Big]\le Cn^{-1/3+2\alpha}$$

Since $\alpha<1/6$ we conclude that

$$\mathbb{E}\Big[\sup_{t\le T}|\epsilon_n^f(t)|\Big]\to 0$$

as $n \to \infty$ as required.

Remark 3.19. *The estimate on* $\mathbb{E}\Big[\sup_{t\leq T} |\epsilon_n^f(t)|\Big]$ *depends on the Lipschitz bound on correlation kernel* $g(x, y)$*. This is precisely the place where our argument brakes down if we allow the environmental noise to be 'white'.*

The only difficulty that we still face is that the compact containment condition, and hence tightness of our sequence of processes, is in the space $\mathcal{M}_\lambda$, and since $\langle \overline{w}_n^2, f\rangle$ cannot be written as an integral with respect to the measure $\overline{M}_n$ or $\overline{M}_n^{\otimes 2}$, convergence of $\langle \overline{w}_n^2, f\rangle$ to $\langle (w^\infty)^2, f\rangle$ is not a simple consequence of the convergence of the sequence of measure-valued evolutions. Evidently, the term in $\overline{w}_n^3$ suffers from the same problem. However, the proof of Theorem 3.10 will be complete, if we can prove the following proposition.

Proposition 3.20. *For* $k = 2, 3$*, we have*

$$\limsup_{n\to\infty} \mathbb{E}\Big[\Big|\langle \overline{w}_n^k(s), f\rangle - \langle (w^\infty)^k(s), f\rangle\Big|\Big] = 0. \tag{3.48}$$

The proof, which we give in Subsection 3.5.3, once again follows Etheridge et al. (2014). It rests on the continuity estimate on the densities $\overline{w}_n$ that we have used above, and integration with respect to a suitable approximate identity.

3.5.2 Uniqueness of solutions for 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 SPDE. The first task is to show that any solution to the martingale problem (3.12) is a weak solution to the SPDE (3.13). Since by Proposition 3.20 the solutions to the limiting martingale problem are continuous, this is a consequence of Theorem 2.85. This then reduces the question to that of the weak uniqueness of the solution to the corresponding SPDE, which we deduce from the pathwise uniqueness for SHE, see Theorem 2.83 and Yamada-Watanabe Theorem for SPDE (see Theorem 2.79).

In one dimension, although any solution to the martingale problem does indeed give a solution to the SPDE, we do not have an analogue of Theorem 2.83. 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 3.3. We note, however, that the proof of our main result would need to be modified to capture this form of environmental noise.

Corollary 3.21. *Solutions to the SPDE (3.13) are unique.*

Proof It is easy to see that all the conditions of Theorem 2.83 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 $C_0(\mathbb{R}^d \times \mathbb{R}^d)$ and so condition (2.21) is satisfied. □

3.5.3 Proof of Proposition 3.20

In this section we shall prove Proposition 3.20 in the case $k = 2$. The case $k = 3$ is entirely analogous.

Let ρ_ϵ be a continuous density in $\mathbb{R}^d$ supported on $B_0(\epsilon)$. Then

$$\begin{aligned}
&\left|\langle \overline{w}_n^2(s), f\rangle - \langle (w^\infty)^2(s), f\rangle\right| \\
&\qquad \le \left|\int_{\mathbb{R}^d} f(x)\overline{w}_n^2(s,x)\mathrm{d}x - \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x)\overline{w}_n(s,x)\overline{w}_n(s,y)\rho_\epsilon(x-y)\mathrm{d}y\mathrm{d}x\right| \\
&\qquad + \left|\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x)w^\infty(s,x)w^\infty(s,y)\rho_\epsilon(x-y)\mathrm{d}y\mathrm{d}x - \int_{\mathbb{R}^d} f(x)(w^\infty(s,x))^2\mathrm{d}x\right| \\
&\qquad\quad + \left|\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x)\overline{w}_n(s,x)\overline{w}_n(s,y)\rho_\epsilon(x-y)\mathrm{d}y\mathrm{d}x\right. \\
&\qquad\qquad \left. - \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x)w^\infty(s,x)w^\infty(s,y)\rho_\epsilon(x-y)\mathrm{d}y\mathrm{d}x\right|.
\end{aligned}$$

From convergence of the process in $\mathcal{M}_\lambda$, we can deduce that the third term converges to zero. We turn our attention to the first term, noticing that the second term can be estimated in the same way. Taking expectations and using Fubini's Theorem,

$$\begin{aligned}
&\mathbb{E}\left[\left|\int_{\mathbb{R}^d} f(x)\overline{w}_n^2(s,x)\mathrm{d}x - \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x)\overline{w}_n(s,x)\overline{w}_n(s,y)\rho_\epsilon(x-y)\mathrm{d}y\mathrm{d}x\right|\right] \\
\le\ & \int_{\mathbb{R}^d}\mathbb{E}\left[\left|f(x)\overline{w}_n(s,x)\int_{\mathbb{R}^d}\overline{w}_n(s,x) - \overline{w}_n(s,y)\rho_\epsilon(x-y)\mathrm{d}y\right|\mathrm{d}x\right] \\
\le\ & \|f\|_\infty \int_{S_f}\int_{B_0(\epsilon)} \mathbb{E}[|\overline{w}_n(s,x) - \overline{w}_n(s,y)|]\rho_\epsilon(x-y)\mathrm{d}y\mathrm{d}x, \qquad (3.49)
\end{aligned}$$

where, as before, S_f denotes the support of the function f.

We now bound the expectation on the right hand side, by proving a continuity estimate for $\overline{w}_n$. In the limit as $n \to \infty$, this results from the smoothing effect of the heat semigroup. We use a well-known trick (Mueller and Tribe (1995)) based on substituting a careful choice of test functions, which approximate the Brownian transition density, into the martingale characterization of the process. We shall focus on $d = 1$, but the proof in higher dimensions follows the same pattern.

First observe that setting $F(x) \equiv x$ in (3.32),

$$\begin{aligned}\mathcal{L}_n^{neu}(\langle w_n, \phi\rangle) = u_n \int_{\mathbb{R}} n\overline{w}_n(y) \int_{B_n(y)} \phi(x)(1 - w_n(x))\mathrm{d}x\mathrm{d}y \\ - u_n \int_{\mathbb{R}} n(1 - \overline{w}_n(y)) \int_{B_n(y)} \phi(x) w_n(x)\mathrm{d}x\mathrm{d}y \\ = nu_n \int\int \phi(x)\mathbf{1}_{\{|x-y|<R_n\}}\Big(\overline{w}_n(y) - w_n(x)\Big)\mathrm{d}x\mathrm{d}y. \quad (3.50)\end{aligned}$$

Writing $\overline{w}_n$ as an integral and using the notation

$$\Lambda(x) = \max(1 - |x|, 0), \qquad \Lambda_R(x) = R\Lambda(x/R),$$

noting that, in $d = 1$, $\mathtt{vol}(B_R(x) \cap B_R(y)) = \Lambda_{2R}(x - y)$, we obtain

$$\mathcal{L}_n^{neu}(\langle w_n, \phi\rangle) = \frac{n^{2/3}\bar{u}}{V_R} \int_{\mathbb{R}} \phi(x)\Big(\int \Lambda_{2R/n^{1/3}}(x - y)\Big(w_n(y) - w_n(x)\Big)\mathrm{d}y\Big)\mathrm{d}x.$$

We now choose a special class of time dependent test functions ϕ_t^n, in such a way that

$$\mathcal{L}_n^{neu}(\langle w_n(s), \phi_{t-s}^n\rangle) = 0.$$

In order to make contact with the notation of Mueller and Tribe (1995), we set $\phi_t^n = \psi_t^{n^{2/3}}$ where ψ_t^n satisfies

$$\frac{\partial}{\partial t}\psi_t^n(x, z) = n\bar{u}\left(\int_{\mathbb{R}}\left(1 - \frac{|x - y|}{2\sqrt{n}}\right)_+\Big(\psi_t^n(y, z) - \psi_t^n(x, z)\Big)\mathrm{d}y\right),$$

with initial condition $\psi_0^n(x;y) = \delta_y(x)$. Writing $\Lambda_{2/\sqrt{n}}^{*k}$ for the k-fold convolution of $\Lambda_{2/\sqrt{n}}$,

$$\psi_t^n(x,z) = \sum_{k=0}^{\infty} e^{-nt} \frac{(nt)^k}{k!} \Lambda_{2/\sqrt{n}}^{*k}(x-z).$$

We shall require the following lemma, which is essentially Lemma 3 of Mueller and Tribe (1995), and can be proved in the same way, using bounds from the local central limit theorem.

Lemma 3.22. *For all x, y, z,*

1. $|\psi_t^n(x,z) - e^{-nt}\delta_z(x)| \le C_T t^{-1/2}$ *for* $t \le T$.
2. $|\psi_t^n(x,y) - \psi_t^n(x,z) - e^{-nt}\delta_y(x) + e^{-nt}\delta_z(x)| \le C(t^{-1}|y-z| + n^{-1}t^{-3/2})$.
3. $\psi_t^n(x,z) \le C_{\lambda,T} e^{-\lambda|x-z|}$ *for* $\lambda > 0$, $t \le T$ *and* $|x-z| \ge 1$.
4. $\|\psi_t^n(x,z) - e^{-nt}\delta_z(x)\|_{L_\lambda^\infty} \le C_{\lambda,T} t^{-1/2} e^{\lambda|z|}$ *for* $\lambda > 0$, $t \le T$.
5. *For* $\lambda > 0$, $t \le T$,

$$\|\psi_t^n(x,y) - \psi_t^n(x,z) - e^{-nt}\delta_y(x) + e^{-nt}\delta_z(x)\|_{L_\lambda^\infty} \le C_{\lambda,T}(t^{-1/2}|y-z|^{1/2} + n^{-1/2}t^{-3/4})e^{\lambda|z|}$$

Now using (3.42) with $F(x) = x$ and (3.44) we find that

$$w_n(t,x) = M_t(\phi_0^n(y,x)) + \int_0^t \int_{\mathbb{R}} \phi_{t-s}^n(z,x) v_n(s,z) \mathrm{d}z\mathrm{d}s + \mathcal{O}(n^{-1/3}),$$

where $M_t(\phi_\cdot^n(\cdot,x))$ is a martingale and $v_n(s,z)$ is uniformly bounded over compact time intervals. We can then write

$$|w_n(t,x) - w_n(t,y)| = \left| M_t(\phi_0^n(x,z)) - M_t(\phi_0^n(y,z)) + \int_0^t \int_{\mathbb{R}} \Big(\phi_{t-s}^n(z,x) - \phi_{t-s}^n(z,y)\Big) v_n(s,z) \mathrm{d}z\mathrm{d}s \right|.$$

We first control the integral on the right. Noting that ϕ^n corresponds to ψ^N in Lemma 3.22 with $N = n^{2/3}$ we write

$$\left|\int_0^t \int_{\mathbb{R}} \Big(\phi^n_{t-s}(z,x) - \phi^n_{t-s}(z,y)\Big) v_n(s,z)\mathrm{d}z\mathrm{d}s\right| \leq I_1 + I_2 + I_3$$

where

$$\begin{aligned}
I_1 &= \left|\int_0^{t-\tau} \int_{\mathbb{R}} \Big(\phi^n_{t-s}(z,x)\,\phi^n_{t-s}(z,y)\Big) v_n(s,z)\mathrm{d}z\mathrm{d}s \right. \\
&\qquad\qquad \left. - \int_0^{t-\tau} e^{-n^{2/3}(t-s)}(v_n(s,x) - v_n(s,y))\mathrm{d}s\right|, \\
I_2 &= \left|\int_{t-\tau}^t \int_{\mathbb{R}} \Big(\phi^n_{t-s}(z,x) - \phi_{t-s}(z,y)\Big) v_n(s,z)\mathrm{d}z\mathrm{d}s - \right. \\
&\qquad\qquad \left. - \int_{t-\tau}^t e^{-n^{2/3}(t-s)}(v_n(s,x) - v_n(s,y))\mathrm{d}s\right|, \\
I_3 &= \left|\int_0^t e^{-n^{2/3}(t-s)}(v_n(s,x) - v_n(s,y))\mathrm{d}s\right|.
\end{aligned}$$

We estimate these three terms separately. First observe that

$$\begin{aligned}
I_1 &= \left|\int_0^{t-\tau} \int_{\mathbb{R}} \Big(\phi^n_{t-s}(z,x) - \phi^n_{t-s}(z,y) + e^{-n^{2/3}t}\delta_x(z) - e^{-n^{2/3}t}\delta_y(z)\Big) v_n(s,z)\mathrm{d}z\mathrm{d}s\right| \\
&\leq \sup_{s\in[0,t-\tau]} \|\phi^n_{t-s}(z,x) - \phi^n_{t-s}(z,y) + e^{-n^{2/3}t}\delta_x(z) - e^{-n^{2/3}t}\delta_y(z)\|_{L^\infty_\lambda} \\
&\qquad\qquad \times \left|\int_0^{t-\tau} \int_{\mathbb{R}} e^{-\lambda z} v_n(s,z)\mathrm{d}z\mathrm{d}s\right| \\
&\leq C_{\lambda,T}\Big(\tau^{-\frac{1}{2}}|x-y|^{\frac{1}{2}} + n^{-\frac{1}{3}}\tau^{-\frac{3}{4}}\Big) e^{\lambda|y|},
\end{aligned}$$

where we have used Part 5 of Lemma 3.22. Analogously,

$$\begin{aligned}
I_2 &= \left|\int_{t-\tau}^t \int_{\mathbb{R}} \Big(\phi^n_{t-s}(z,x) - \phi^n_{t-s}(z,y) + e^{-n^{2/3}t}\delta_x(z) - e^{-n2/3t}\delta_y(z)\Big) v_n(s,z)\mathrm{d}z\mathrm{d}s\right| \\
&\leq C_\lambda \int_{t-\tau}^t \|\phi^n_{t-s}(z,x) - \phi^n_{t-s}(z,y) + e^{-n^{2/3}t}\delta_x(z) - e^{-n^{2/3}t}\delta_y(z)\|_{L^\infty_\lambda}\mathrm{d}s \\
&\leq C_{\lambda,t}\tau^{\frac{1}{2}} e^{\lambda|y|},
\end{aligned}$$

where we have used Part 4 of Lemma 3.22. Finally, I_3 can be bounded by

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

Combining the estimates above we conclude that

$$\left|\int_0^t \int_{\mathbb{R}} \phi^n_{t-s}(z,x) - \phi^n_{t-s}(z,y) v_n(s,z) \mathrm{d}z \mathrm{d}s\right| \leq C\Big(|x-y|^{1/4} + n^{-1/12}\Big) e^{\lambda|x|}. \tag{3.51}$$

We turn our attention to the difference of martingale terms. We use the Burkholder-Davies-Gundy inequality in the form:

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

The first term can be controlled in the same way as above, by writing it as an integral of $\phi^n_\cdot(\cdot,x) - \phi^n_\cdot(\cdot,y)$ times a bounded function and the jumps of the martingale are of size $\mathcal{O}(n^{-1/3})$, so combining with (3.51) leads to

$$\mathbb{E}[|\overline{w}_n(s,x) - \overline{w}_n(s,y)|] \leq C_{\lambda,t}\Big(\Big(n^{-\frac{1}{3}} + |x-y|^{\frac{1}{4}} + n^{-\frac{1}{12}}\Big) e^{\lambda|x|}\Big).$$

Taking into account that ρ_ϵ is a probability density supported in $B_0(\epsilon)$ and that the support of the test function f is compact, we conclude that (3.49) can be bounded by

$$\mathbb{E}\left[\left|\int_{\mathbb{R}^d} f(x)\overline{w}_n^2(s,x)\mathrm{d}x - \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x)\overline{w}_n(s,x)\overline{w}_n(s,y)\rho_\epsilon(x-y)\mathrm{d}y\mathrm{d}x\right|\right] \leq C_{\lambda,t}\Big(n^{-\frac{1}{3}} + \epsilon^{\frac{1}{4}} + n^{-\frac{1}{12}}\Big)$$

Noting once again that the remaining term can be bounded in the same way, we let $n\to\infty$ to obtain

$$\lim_{n\to\infty} \mathbb{E}\Big[\Big|\langle \overline{w}_n^k(s), f\rangle - \langle (w^\infty)^k(s), f\rangle\Big|\Big] \leq \epsilon^{\frac{1}{4}},$$

which, since the choice of ϵ was arbitrary, completes the proof of Proposition 3.20 .

Chapter 4

Lookdown constructions

4.1 Introduction

Most of the models discussed in this thesis belong to the class of measure-valued processes. Many of them arise naturally as infinite limits of empirical measures of particle systems. In this chapter we describe an approach in which, in sharp contrast to models coming from diffusion approximations, distinct particles are still present in the limit. This type of construction was introduced in Donnelly and Kurtz (1996, 1999a). It is usually called a lookdown construction (or lookdown representation). This approach has turned out to be particularly fruitful in applications to population models. In that setting, typically, each individual in the population is assigned a 'level' (taking values in either the integers or the reals). Levels usually carry information about genealogical relations between individuals. The origin of the name 'lookdown' is that individuals usually determine their parents by 'looking down' at the subpopulation with levels lower than their own.

From a practical perspective, one of the most useful properties of lookdown constructions is that in the passage from the individual based models to their high density limit, or continuous approximation, the genealogies are preserved, even though the individuals have disappeared from the picture. To see why this is important we refer to the examples of systems of individual based models which are approximated by the same

diffusion processes even though the distribution of their genealogies differ significantly obtained in Taylor (2009).

Let us briefly describe two simple examples. The first one is taken from the original work of Donnelly and Kurtz (1996). Consider a population of N individuals. Each of the individuals is assigned a level l_i from $\{1, 2, \ldots, N\}$. The initial assignment of levels is performed by drawing randomly from a uniform distribution over all possible assignments.

The evolution of the population is very simple. Each pair of individuals is assigned an independent Poisson process $\pi_{i,j}$ with rate τ. This Poisson process determines the times of replacement events. Whenever a replacement event occurs, the values of the levels l_i and l_j are compared. If the values of the levels are different, the type of the individual with the higher level is replaced by the type of the individual with the lower one. This means that during each reproduction event the offspring (the individual whose type was replaced) 'looks down' to find the parent. Let $\Phi^N_{i,j}$, acting on a vector $\overline{x}$ of length N, be defined as the replacement of the jth position of the vector with the ith one, that is

$$\Phi^N_{i,j}[(x_1, x_2, \ldots x_i \ldots, x_j, \ldots x_N)] = (x_1, x_2, \ldots x_i \ldots, x_i, \ldots x_N).$$

Then the generator of the process described above is given by

$$A^N f(\overline{x}) = \sum_{i \neq j} \tau\{f(\Phi_{i,j}(\overline{x})) - f(\overline{x})\}. \tag{4.1}$$

This generator does not resemble a sensible population model. However, if we average over the initial distribution of the levels the resulting model is just a classical Moran model, Moran (1958). Let us describe the averaging procedure in a little more detail.

Consider a state space of types E and the space of all probability measures $\mathcal{P}(E)$. For each $\overline{x} \in E^N$ consider the mapping $z_N : E^N \to \mathcal{P}(E)$ given by

$$z_N(\overline{x}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}.$$

Observe that if $\overline{x}_t$ is a solution to the martingale problem associated with generator (4.1), z_n is just the empirical distribution of our process. For a continuous bounded function $f : E^N \to \mathbb{R}$, consider the map α given by

$$\alpha f = \frac{1}{N!} \sum_{\sigma} f(x_{\sigma(1)}, \ldots, x_{\sigma(N)}),$$

which is just an average over all possible permutations of the set $\{1, 2, \ldots, N\}$. The role of the map α is to average over all possible initial assignments of levels. We observe that if we average the generator (4.1) it leads to

$$\alpha A^N f(z_n) = \frac{1}{2} \sum_{1 \leq i \neq j \leq N} \tau \Big\{ \alpha f(z_N - \frac{1}{N}(\delta_{x_i} - \delta_{x_j})) - \alpha f(z_N) \Big\} = C_N \alpha f,$$

where C_N is the generator of the classical Moran model acting on αf. The Markov Mapping Theorem implies that for any solution $\tilde{Y}_N$ of the martingale problem associated with generator C_N there exists a solution X_N of the martingale problem associated with generator (4.1) such that $\tilde{Y}_N$ and X_N have the same distribution. So what we have really shown is that our initial process, described by the generator (4.1), is just the usual Moran model with an unusual labelling of the particles.

The initial constructions of Donnelly and Kurtz (1996, 1999a,b) rely on properties of exchangeable sequences. For example, the proof of equivalence of the Moran model and its lookdown representation in Donnelly and Kurtz (1996) is based on an explicit construction and a filtering argument. The corresponding proof in Donnelly and Kurtz (1999b) is based on a coupling argument. Both are significantly more involved than the approach we have just presented.

We move on to our second example. Let us consider a simple model in which the size of the population decreases by one in each event. Consider a system of n particles. Each particle, independently, dies at rate b. The generator of this simple process, which we call the pure death process, for $\widehat{f} \in C_c^\infty$ can be written as

$$A\widehat{f}(n) = b(\widehat{f}(n-1) - \widehat{f}(n)). \tag{4.2}$$

We could define the same process in a different way. Once again we start with a system of n particles. Fix $\lambda \in \mathbb{R}_+$. Each particle $i \in \{0, 1, \dots, n\}$ is assigned a number $l_i \in [0, \lambda)$. We refer to the number l_i as the level of the particle i. The evolution of l_i in time is ruled by an ordinary differential equation of the form

$$\partial_t l_i(t) = b l_i(t).$$

We assume that the particle is present in the system as long as its level l_i is below λ. When the level of the particle hits λ, the particle dies and is removed from the system.

So far, the behaviour of this system is purely deterministic. In order to bring in the randomness (and to make the connection to the process defined by the generator (4.2)), we allow the the initial level of each of the particles to be drawn randomly from the uniform distribution on $[0, \lambda)$, independently of all the other levels.

Since all the ingredients of our construction are now in place, we may describe the process more formally. Our system is described by a pair $(N(t), L(t))$, where

$$N(t) = \#\{i : l_i < \lambda\}, \quad L(t) = (l_1, l_2, \dots, l_i \dots),$$

that is, $N(t)$ is the number of particles in the system which are alive at time t and $L(t)$ is the collection of their levels.

Consider a set of test functions of the form

$$f(l, n) = \prod_{i=1}^{n} g(l_i), \tag{4.3}$$

where the function g is twice continuously differentiable. The generator of the lookdown representation of the pure death process is given by

$$\partial_t f(L, N) = A f(l, n),$$

where the operator A can be evaluated using the product rule for differentiation as follows:

$$Af(l,n) = \sum_{i}^{n} \left(bl_i g'(l_i) \prod_{j \neq i} g(l_j) \right) = f(l,n) \sum_{i}^{n} \left(bl_i \frac{g'(l_i)}{g(l_i)} \right), \tag{4.4}$$

where we have also used the definition of the function f.

Remark 4.1. *The last expression in formula* (4.4) *is usually the most convenient to use in calculations. We observe that it is always well-defined, since whenever* $g(l_i) = 0$*, we interpret it as the middle sum in* (4.4)*.*

Now we are in position to make the connection between the lookdown representation of the pure death process and the process defined by the generator (4.2). As was the case for the Moran model, our aim is to average out the distribution of the levels and then, to make the connection rigorous, apply the Markov Mapping Theorem. Let $\alpha_{n,\lambda}(n, \mathrm{d}l)$ be the joint distribution of n i.i.d. uniform random variables on $[0, \lambda]$. The average over one level with respect to a test function g, which we denote by $e^{-\mathcal{I}_\lambda(g)}$, is given by

$$e^{-\mathcal{I}_\lambda(g)} = \frac{1}{\lambda} \int_0^\lambda g(z) \mathrm{d}z.$$

Therefore, for a test function f of the form (4.3), the averaged test function $\widehat{f}$ is given by

$$\widehat{f}(n) = \int f(l,n) \mathrm{d}\alpha_\lambda(n, \mathrm{d}l) = e^{-n\mathcal{I}_\lambda(g)}.$$

We evaluate the averaged generator of the lookdown representation of the pure death process (4.4) by computing

$$\begin{aligned} \int Af(l,n) \mathrm{d}\alpha_\lambda(n, \mathrm{d}l) &= \sum_{i=1}^{n} e^{-(n-1)\mathcal{I}_\lambda(g)} b \frac{1}{\lambda} \int_0^\lambda z g'(z) \mathrm{d}z \\ &= bne^{-(n-1)\mathcal{I}_\lambda(g)} \frac{1}{\lambda} \int_0^\lambda (g(\lambda) - g(z)) \mathrm{d}z \\ &= bne^{-(n-1)\mathcal{I}_\lambda(g)} (1 - e^{-\mathcal{I}_\lambda(g)}) = bn(\widehat{f}(n-1) - \widehat{f}(n)) = C\widehat{f}(n). \end{aligned}$$

Clearly C is the generator of the pure death process (4.2) acting on $\widehat{f}$. To formalize the connection we should check all the assumptions of the Markov Mapping Theorem. The generator of the process (4.2) is the generator C from the statement of the theorem, the generator A is given by (4.4). It is sufficient to consider the map ψ given by $\psi(l,n) = n+1$. The moment condition (1.11) is satisfied as long as $\mathbb{E}N_0 < \infty$. The generator is bounded on the set of test functions (4.3). Therefore by the Markov Mapping Theorem there exists a map γ such that if X is a solution to the martingale problem specified by generator (4.2), and Y is a solution to the martingale problem specified by the generator (4.4) then $\gamma(Y)$ has the same distribution as X.

The rest of the chapter is organised as follows. In Section 4.2 we present a new lookdown construction of the SuperBrownian motion in a random environment. In Section 4.3 we recall two constructions of the neutral Spatial Lambda-Fleming-Viot model. The first one was introduced in Véber and Wakolbinger (2015) and the second one was introduced by Etheridge and Kurtz (2018). Section 4.4, and 4.5 are devoted to the main results of this chapter. In Section 4.4, we use a lookdown construction of the Lambda-Fleming-Viot model with selection in a fluctuating environment inspired by that of the neutral model in Section 4.1.3, Etheridge and Kurtz (2018) to show that a scaling which captures the behaviour of a 'rare' type converges to the Feller Diffusion in random environment. In Section 4.5 we show how to extend the results of Section 4.4 to show an analogous result for the Spatial Lambda-Fleming-Viot model selection in a fluctuating environment.

4.2 SuperBrownian motion in a random environment

In this section we present a precise definition of the superBrownian motion in a random environment. We begin by defining Branching Brownian Motion in a random environment and recalling the original definition of the corresponding superprocess from Mytnik (1996b). Then we describe a lookdown construction for both of these models based on the ideas in Kurtz and Rodrigues (2011).

4.2.1 Definitions

Branching Brownian Motion in a random environment (BBMRE) can be described as follows. Imagine a collection of particles on $\mathbb{R}^d$. Each particle moves according to independent standard Brownian motions. Each particle, if alive, gives birth to one new particle at a time, at rate a. The initial offspring location is the same as that of the parent. After the birth, the offspring moves and reproduces independently of all other particles. The particles die at instantaneous rate $a - \zeta_t(x)b$, where $\zeta_t(x)$, taking values in $\{-1, 1\}$ as before, models the random environment and x is the current location of the particle. We assume that $a > b > 0$. If $\zeta_t(x)$ is positive, the particle is less likely to die, if it is negative, it is more likely to do so. The evolution of the environment and particles are independent. We give a more formal definition below.

We begin with a description of the environment, which is used for all models in this section. Our environment is modelled through a simple random field.

Definition 4.2. *Let Π^{env} be a Poisson process with intensity E, dictating the times of the changes in the environment. Let $q(x, y)$ be an element of $C_0\Big(\mathbb{R}^d \times \mathbb{R}^d\Big)$ (continuous functions vanishing at infinity) and let $\{\xi^{(m)}(\cdot)\}_{m\geq 0}$ be a family of identically distributed random fields on $\mathbb{R}^d$ such that*

$$\begin{aligned}\mathbb{P}\Big[\xi^{(m)}(x) = -1\Big] =& \frac{1}{2} = \mathbb{P}\Big[\xi^{(m)}(x) = +1\Big],\\ \mathbb{E}\Big[\xi^{(m)}(x)\xi^{(m)}(y)\Big] =& q(x, y).\end{aligned}$$

Set $\tau_0 = 0$ and write $\{\tau_m\}_{m\geq 1}$ for the points in Π^{env} and define

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

For the construction of the random field in Definition 4.2 we refer to Ma (2009), especially their Example 1. The arguments presented there require only a minor adaptation. We observe that the generator of the process describing the evolution of the environment, A^{env}, is given by

$$A^{env} f(\zeta) = \mathbb{E}_\pi[f(\zeta)] - f(\zeta), \tag{4.5}$$

where π is the stationary distribution of the random field ζ.

Since the exact labelling of our particles is not important, we identify the particle component of the process with a counting measure, that is for a vector $\overline{x} = (x_1, \dots x_n)$

$$\mu_{\overline{x}} = \sum_i \delta_{x_i} \quad x_i \in \mathbb{R}^d.$$

We are now ready to state the definition.

Definition 4.3 (Branching Brownian motion in a random environment (BBMRE))**.** *Branching Brownian motion in the random environment ζ is the stochastic process taking values in purely atomic measures on $\{-1, 1\} \times \mathbb{R}^d$ whose evolution consists of four ingredients.*

1. ***Spatial motion*** *The location of each particle, x_i, evolves according to a standard Brownian motion, independently of all other particles.*

2. ***Birth events*** *At exponential rate a (independent for each particle), a particle gives birth to a new particle (a new particle is added to the system). The location of the offspring is the same as the location of their parent. The behaviour of the new particle after the birth is independent of all other particles.*

3. ***Death events*** *Each particle dies (is removed from the system) at instantaneous rate $a - \zeta_t(x_i)b$, where x_i is its location.*

4. ***Environment changes*** *The environment evolves as described in Definition 4.2.*

Alternatively, we may define the BBMRE by the means of the generator. Recall that we identify our process with a counting measure on $\mathbb{R}^d$. For a counting measure $\mu = \sum_i \delta_{x_i}$, define

$$f(\mu) = \sum_i f(x_i).$$

Let $f(\zeta, \mu) = f_0(\zeta) f_1(\mu)$ be a function such that $f_1 \in C_c(\mathbb{R}^d)$, that is f_1 is a continuous function with a compact support. We define the generator of the BBMRE as

$$\mathcal{L} f(\zeta, \mu)$$

$$= f_1(\mu)A^{env}f_0(\zeta) + f_0(\zeta)\Bigg(\sum_{i=1}^{n} Bf_1(\mu) + \sum_i a(f_1(\mu_{b(x|x_i)}) - f_1(\mu)) + \sum_i \left(a - \zeta(x_i)b(f_1(\mu_{d(x|x_i)}) - f_1(\mu)\right)\Bigg),$$

where $\mu_{b(x|x_i)}$ denotes the addition of a particle at location x_i, and $\mu_{d(x|x_i)}$ denotes a removal of the particle at location x_i.

It is well known that the high density limit for the Branching Brownian Motion gives rise to a SuperBrownian motion (see e.g. Etheridge (2000)). An analogous result (under certain scaling of the environment) was established by Mytnik (1996b) for branching random walk in a random environment. The limiting object is SuperBrownian motion in a random environment (SBMRE).

Definition 4.4 (SuperBrownian motion in a random environment (SBMRE))**.** *Let $q(x,y) \in C_0(\mathbb{R}^d \times \mathbb{R}^d)$ be a covariance function. The superBrownian motion in a random environment is the (unique) process, taking values in $\mathcal{M}_f(\mathbb{R}^d)$ (that is, the space of finite measures on $\mathbb{R}^d$), for which, for all $\phi \in \mathcal{D}(\Delta)$,*

$$X_t(\phi) = X_t(\phi) - X_0(\phi) - \int_0^t \frac{1}{2} X_s(\Delta\phi)\mathrm{d}s \tag{4.6}$$

is a square-integrable martingale with quadratic variation given by

$$\langle X(\phi)\rangle_t = \int_0^t X_s(\phi^2)\mathrm{d}s + \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} q(x,y)\phi(x)\phi(y)X_s(\mathrm{d}x)X_s(\mathrm{d}y)\mathrm{d}s.$$

An equivalent characterisation of the SBMRE can be given in terms of the generator, see Theorem 4.8 in Mytnik (1996b). Namely, for $f \in \bar{C}^2(R_+), \phi \in \mathcal{D}(\Delta)$ and $q(x,y)$ as in Definition 4.4, the generator is given by

$$\mathcal{L}f(\mu(\phi)) = f'(\mu(\phi))\mu(\Delta\phi) + \frac{1}{2}f''(\mu(\phi))\Bigg(\mu(\phi^2) + \int_{\mathbb{R}^d\times\mathbb{R}^d} q(x,y)\phi(x)\phi(y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\mathrm{d}s\Bigg).$$

We would like to point out that the process which is obtained as a limiting behaviour of the scalled SLFV in a random environment differs from the one from Definition 4.4 by a presence of a drift term. Namely, for $f \in \bar{C}^2(R_+), \phi \in \mathcal{D}(\Delta)$ and $q(x,y)$ as in

Definition 4.4, we are considering the process with the generator

$$\mathcal{L}f(\mu(\phi)) = f'(\mu(\phi))[\mu(\Delta\phi) + b\mu(\phi))] + \frac{1}{2}f''(\mu(\phi))\Big(\mu(\phi^2) + \int_{\mathbb{R}^d\times\mathbb{R}^d} q(x,y)\phi(x)\phi(y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\mathrm{d}s\Big). \quad (4.7)$$

Remark 4.5. *Uniqueness of solutions to the martingale problem of Definition 4.4 is not immediately clear. It was established by Mytnik (1996b) using a novel approximate duality technique, and later re-proved by the means of the log-Laplace transform by Crisan (2004). A uniqueness result for the model in the presence of additional drift is a simple consequence of Dawson's Girsanov Theorem, see Dawson (1978) and Etheridge (2000), Chapter 7.*

Remark 4.6. *A model similar to that in Mytnik (1996b) was studied in Sturm (2003). The main difference between the two is in the behaviour of the environment. For the limiting model in Sturm (2003),* (4.6) *is again a martingale, but with quadratic variation of the form*

$$\langle X(\phi)\rangle_t = \int_0^t \int_{\mathbb{R}^d\times\mathbb{R}^d} q(x,y)\phi(x)\phi(y)X_s(\mathrm{d}x)X_s(\mathrm{d}y)\mathrm{d}s.$$

In this case, the density of the process can be described as a solution to an SPDE in all dimensions, whereas the SBMRE has a density only in dimension one and the analogous SPDE has no solution in dimensions $d \geq 2$.

Remark 4.7. *An alternative construction of the SBMRE has been suggested by Nakashima (2015). The construction in this paper is based on the model introduced by Birkner et al. (2005).*

4.2.2 Lookdown construction for BBMRE and SBMRE

We now describe a new construction of the SBMRE, inspired by the constructions of Kurtz and Rodrigues (2011). As a by-product of this construction, we provide a lookdown construction for the BBMRE. A precise statement of the results is given in Theorem 4.9 and Theorem 4.10. We will use the lookdown representation of SBMRE in Section 4.5 to describe the behaviour of the 'rare' type (by which we mean a new

mutation establishing in the population) in the Spatial Lambda-Fleming-Viot model with fluctuating selection.

In order to motivate what follows, let us informally describe the construction of SBMRE from Mytnik (1996b). The SBMRE is constructed via a series of approximations. Consider a sequence of mean 0 random fields ζ_k, taking values in $\{-1, 1\}$, with correlation as in Definition 4.2 (Mytnik (1996b) considered a more general class of random fields, but this one is sufficient for our purposes). At stage n of the approximation, we start with a population with size of order n. Over each time interval $(k/n, (k+1)/n), k \in \mathbb{N}$, each individual, if alive, moves independently of all the others, according to a standard Brownian motion. At times k/n, each individual either splits into two with probability $1/2+\zeta_k(x_i)/\sqrt{n}$, or dies with probability $1/2-\zeta_k(x_i)/\sqrt{n}$. The state of the environment is resampled with each reproduction event. If for each set B we define

$$X_t^n(B) = \frac{\text{number of particles in } B \text{ alive at time } t}{n},$$

and assume that X_0^n converges to some X_0, then passing to the limit as n tends to infinity the process X^n converges to a SBMRE with initial condition X_0. Intuitively, this procedure corresponds to increasing the rate of the branching events in BBMRE by n, while scaling down the impact of the environment by $\sqrt{n}$. Observe that Mytnik (1996b) considers a model with non-overlapping generations. This is for purely technical reasons.

We now move to a description of a lookdown construction for BBMRE (with overlapping generations). Fix $\lambda \in \mathbb{R}$. Consider a system of particles in the geographical space $\mathbb{R}^d$. Each particle moves independently, according to a standard Brownian motion. Each particle i is assigned a level, l_i. The level takes values in $[0, \lambda]$.

In order to take environmental fluctuations into account we use the process $\zeta_t(x) \in \{-1, 1\}$, introduced in Definition 4.2, to model the environment. The branching rate of the particles depends on both their level and the state of the environment. We assume that every particle gives birth at instantaneous rate $2a(\lambda - l_i(t))$, with the location of offspring being the same as the location of the parent. The initial level of the offspring is distributed uniformly on the interval $[l_i, \lambda]$. The levels of the particles evolve according

to an ODE with a random coefficient

$$\frac{\mathrm{d}l_i}{\mathrm{d}t} = al_i^2 - \zeta(x_i)\sqrt{\lambda}bl_i.$$

The particle dies when its level reaches λ. We are now ready to define the process of interest.

Definition 4.8 (Lookdown representation of Branching Brownian motion in a random environment)**.** *The lookdown representation of BBMRE is the process taking values in* $(\mathbb{R}^d \times \mathbb{R})^\infty \times \{-1, 1\}$ *with dynamics specified by four components.*

1. ***Spatial motion*** *The spatial location of each particle,* $x_i \in \mathbb{R}^d$*, evolves according to a standard Brownian motion with generator* $B = \frac{1}{2}\Delta$*.*

2. ***Birth events*** *Each particle gives birth at instantaneous rate* $2a(\lambda - l_i(t))$*. The location of the offspring is the same as the location of their parent. If* l^* *denotes the level of the parent, the level of the offspring is chosen uniformly at random from* $[l^*, \lambda]$*.*

3. ***Level movement*** *The level of each of the particles evolves according to the equation*

$$\frac{\mathrm{d}l_i}{\mathrm{d}t} = al_i^2 - \zeta(x_i)\sqrt{\lambda}bl_i.$$

4. ***Environment changes*** *The environment evolves as described in Definition 4.2.*

A formal definition, via the generator of the process, is given in (4.10). It is convenient to identify the process with the counting measure

$$\mu_{\overline{x},\overline{l}} = \sum_i \delta_{x_i,l_i}.$$

We now state the first result of this section.

Theorem 4.9 (Lookdown construction for Branching Brownian Motion in random environment)**.** *Let* μ_0 *be the measure associated with the initial state of the population. Let* $\gamma : \mathcal{N}(\mathbb{R}^d \times [0, \lambda)) \to \mathcal{M}(\mathbb{R}^d)$ *be given by*

$$\gamma\left(\sum_i \delta_{x_i,l_i}\right) = \frac{1}{\lambda}\sum_{l_i} \delta_{x_i}$$

The lookdown process of Definition 4.8 corresponds to BBMRE, in the sense that if $\eta(\overline{x},\bar{l})$ *is a solution to the martingale problem for the process defined in Definition 4.8, then* $\gamma\Big(\eta(\overline{x},\bar{l})\Big)$ *is the solution to the martingale problem for the process of Definition 4.3.*

Our next aim is to write down the generator of the process obtained by passing to the limit $\lambda \to \infty$, which would correspond to passing to the limit with n passing to infinity in the sequence of approximations described at the beginning of this section. In order to formulate the result for the limiting process, we need to consider a special class of test functions of the form

$$f(\zeta, x, l) = f_0(\zeta) f_1(x, l) = f_0(\zeta) \prod_i g(x_i, l_i), \tag{4.8}$$

with the additional requirement

$$g(x_i, l_i) = 1 \text{ for } l_i > \lambda_g,$$

which ensures that we ignore all individuals with levels above λ_g. We are now in a position to state the main result of this section.

Theorem 4.10 (Lookdown construction for superBrownian motion in random environment)**.** *Consider the process with generator given by*

$$\begin{aligned} A_\infty(f_1) = f_1(x,l)\sum_i \frac{Bg(x_i,l_i)}{g(x_i,l_i)} \\ + f_1(x,l)\sum_i 2a\int_{l_i}^{\infty}(g(x_i,v)-1)\mathrm{d}v + f_1(x,l)\sum_i (al_i^2 - b^2 l_i)\frac{\partial_l g(x_i,l_i)}{g(x_i,l_i)} \\ - f_1(x,l)\sum_i b^2 l_i\left(\sum_{j\neq i} q(x_i,x_j) l_j \frac{\partial_l g(x_i,l_i)\partial_l g(x_j,l_j)}{g(x_i,l_j)g(x_i,l_i)} + \frac{l_i \partial_l^2 g(x_i,l_i)}{g(x_i,l_i)}\right). \end{aligned} \tag{4.9}$$

Let μ_0 be the measure associated with the initial state of the population in the limit. Define the map $\gamma : \mathcal{N}(\mathbb{R}^d \times [0,\infty)) \to \mathcal{M}(\mathbb{R}^d)$ by

$$\gamma\left(\sum_i \delta_{x_i,l_i}\right) = \begin{cases} \lim_{\lambda\to\infty} \frac{1}{\lambda} \sum_{l_i \le \lambda} \delta_{x_i} & \text{if the limit exists,} \\ \mu_0 & \text{otherwise .} \end{cases}$$

The process described by the limiting generator (4.9) *corresponds to the SBMRE of Definition 4.4, in the sense that if $\eta(\overline{x}, \overline{l})$ is a solution to the martingale problem for the process described by the limiting generator* (4.9)*, then $\gamma(\eta(\overline{x}, \overline{l}))$ is a solution to the martingale problem for the process in Definition 4.4.*

We now wish to specify the generator for the process in Definition 4.8. In general, we consider the set of test functions of the form (4.8). To make the functional setup more precise, we need to borrow the following condition from Kurtz and Rodrigues (2011), which guarantees that the assumptions of the Markov Mapping Theorem are satisfied.

Condition 4.11 (Based on Kurtz and Rodrigues (2011), Condition 3.1)**.** *We assume that the following conditions on the operator B, the coefficients a, b and the test functions g are satisfied.*

1. *The operator B is defined on a subset of the space of bounded continuous functions, and its domain is closed under multiplication and separating.*

2. *The test functions are of the form* (4.8)*, where*

$$g(x,l) = \prod_{j=1}^{m} (1 - g_1^j(x) g_2^j(l))$$

and $g_1^j \in \mathcal{D}(B)$ and g_2^j are twice differentiable with support in $[0,\lambda]$. Moreover,

$$0 \le g_1^j g_2^j < \rho_g < 1.$$

3. *There exists a continuous, non-negative function ψ_B such that for every test function g, and for every $x \in \mathbb{R}^d$*

$$\sup_l |Bg(x,l)| \leq c_g \psi_B(x)$$

for some constant c_g which depends only on the function g.

4. *The following bound holds for the test functions:*

$$\int_0^\infty |g(x,l) - 1| \mathrm{d}l + \sup_l \{l + l^2\} \partial_l g(x,l) \leq c_g \psi_B(x)$$

for some constant c_g' which depends only on the function g.

5. $a > 0$, $\lambda a - \sqrt{\lambda} b > 0$.

Remark 4.12. *We notice that the conditions for the generator of the motion B are satisfied by the Laplacian, the generator of the Brownian motion. This is the only generator that we consider. We provide the general construction for the sake of completeness and to highlight potential extensions of this work.*

The generator A_λ of the process of Definition 4.8 can be written as

$$\begin{aligned} A_\lambda f(\zeta,x,l) = f_1(x,l) A_\lambda^{env} f_0(\zeta) + f(\zeta,x,l) \sum_i \frac{Bg(x_i,l_i)}{g(x_i,l_i)} \\ + f(\zeta,x,l) \sum_i 2a \int_{l_i}^\lambda (g(x_i,v) - 1) \mathrm{d}v \\ + f(\zeta,x,l) \sum_i (a l_i^2 - \sqrt{\lambda} \zeta(x_i) b l_i) \frac{\partial_{l_i} g(x_i,l_i)}{g(x_i,l_i)}. \end{aligned} \quad (4.10)$$

Proof[of Theorem 4.9] Before we can proceed, additional objects need to be defined. Let $\alpha_\lambda(n,l)$ be the joint distribution of n i.i.d. uniformly distributed random variables on $[0,\lambda]$. For a test function $f_1(x_i,l_i)$, we define the projection onto type space $\hat{f}$ as

$$\hat{f}(\mu) = \prod_i \hat{g}(x_i) = e^{-\sum_i \mathfrak{I}(g(x_i))},$$

where the average for a single level is defined as

$$e^{-\mathcal{J}(g(x_i))} = \hat{g}(x_i) = \frac{1}{\lambda}\int_0^\lambda g(x_i, z)\mathrm{d}z.$$

To calculate the generator of the projected model (the generator averaged over the distribution of the levels), we need to evaluate

$$\int A_\lambda f(\zeta, x, l)\alpha_\lambda(\mathrm{d}l).$$

Let us integrate the four terms appearing in A_λ separately. We begin with the two terms which are least involved - the movement of particles and the environment. Since both of those terms do not depend on the levels, integrals with respect to them do not alter our projections, namely

$$\int f(\zeta, x, l)\sum_i \frac{Bg(x_i, l_i)}{g(x_i, l_i)}\alpha_\lambda(\mathrm{d}l) = \sum_{i=1}^n B\hat{f}(\zeta, \mu). \tag{4.11}$$

Analogously,

$$\int \lambda f_1(x, l)A_\lambda^{env} f_0(\zeta, x)\alpha_\lambda(n, \mathrm{d}l) = \lambda\hat{f}_1(\mu)A_\lambda^{env} f_0(\zeta, \mu). \tag{4.12}$$

In order to evaluate terms describing births and movement of the levels, which both do depend on the exact value of the level, it is convenient to note that (here we follow the calculation on p. 492 in Kurtz and Rodrigues (2011))

$$\lambda^{-1}2a\int_0^\lambda g(x, z)\int_z^\lambda (g(x, v) - 1)\mathrm{d}v\mathrm{d}z = a\lambda e^{-\mathcal{J}_g} - 2a\lambda^{-1}\int_0^\lambda g(x, z)(\lambda - z)\mathrm{d}z, \tag{4.13}$$

where we have used Fubini's Theorem, and

$$\begin{aligned}\lambda^{-1}\int_0^\lambda (az^2 - \zeta bz)g'(x, z)\mathrm{d}z &= -\lambda^{-1}\int_0^\lambda (2az - \zeta b)(g(z) - 1)\mathrm{d}z \\ &= \lambda^{-1}2a\int_0^\lambda zg(x, z)\mathrm{d}z + a\lambda + b(e^{-\mathcal{J}_g} - 1),\end{aligned} \tag{4.14}$$

where we have integrated by parts. It will also be useful to describe the changes in our system due to deaths and births. Whenever a birth event occurs, the new individual

is located at the same place as the parent. If a death occurs, the individual is just removed from the system. Therefore, if we denote the new collection of particles after a birth at location y by $(b(\overline{x}|y))$ and the new collection of particles after a death at location x_j by $d(\overline{x}|x_j)$, we see that

$$\mu_{b(\overline{x}|y)} = \delta_y + \sum_{i=1}^{n} \delta_{x_i}, \quad \mu_{d(\overline{x}|x_j)} = -\delta_{x_j} + \sum_{i=1}^{n} \delta_{x_i}.$$

Armed with these observations and identities (4.13), (4.14) we proceed to evaluate the remaining terms. A simple calculation shows that

$$\begin{aligned}
&\int f(\zeta,x,l)\Big\{\sum_i^n 2a\int_{l_i}^{\lambda}(g(x_i,v)-1)\mathrm{d}v + \sum_{i=1}^{n}(al_i^2-\sqrt{\lambda}\zeta(x_i)bl_i)\frac{\partial_{l_i}g(x_i,l_i)}{g(x_i,l_i)}\Big\}\alpha_\lambda(\mathrm{d}l)\\
&\qquad = \sum_j e^{\sum_{i\neq j}\jmath_{g(x_i)}}\Big\{2a\lambda e^{-\jmath_{g(x_j)}} + 2a\lambda^{-1}\int_0^{\lambda} g(x_j,z)\mathrm{d}z\\
&\qquad\quad -2a\lambda^{-1}\int_0^{\lambda} g(x_j,z)\mathrm{d}z + a\lambda + \sqrt{\lambda}b\zeta(x_j)(e^{\jmath_{g(x_j)}}-1)\Big\}\\
&= a\lambda e^{\sum_i \jmath_{g(x_i)}}\sum_j(e^{-\jmath_{g(x_j)}}-1) + \sum_j(\lambda a-\sqrt{\lambda}b\zeta(x_j))e^{\sum_{i\neq j}\jmath_{g(x_i)}}(1-e^{\jmath_{g(x_j)}})\\
&= \sum_i \lambda a(\hat{f}(\mu_{b(x|x_i)})-\hat{f}(\mu)) + \sum_i(\lambda a-\sqrt{\lambda}\zeta(x_i)b)(\hat{f}(\mu_{d(x|x_i)})-\hat{f}(\mu)). \quad (4.15)
\end{aligned}$$

Combining (4.11), (4.12) and (4.15) we have established that the projected generator can be written as

$$\begin{aligned}
\mathcal{L}_\lambda\hat{f}(\zeta,\mu) = \lambda\hat{f}_1(\mu)A_\lambda^{env}f_0(\xi) + \sum_i B_{x_i}\hat{f}(\mu) + \sum_i \lambda a(\hat{f}(\mu_{b(x|x_i)})-\hat{f}(\mu))\\
+\sum_i(\lambda a-\zeta(x_i)\sqrt{\lambda}b)(\hat{f}(\mu_{d(x|x_i)})-\hat{f}(\mu)),
\end{aligned}$$

which is the generator of the BBMRE with birth rate λa and death rate $(\lambda a - \sqrt{\lambda}\zeta b)$, as claimed.

Now we only need to check that all assumptions of the Markov Mapping Theorem are satisfied. Fortunately our Condition 4.11 has been imposed to guarantee just that. The map $\gamma : \mathcal{N}(\mathbb{R}^d \times [0,\lambda)) \to \mathcal{M}(\mathbb{R}^d)$ (mapping counting measures to measures on $\mathbb{R}^d$)

is given by

$$\gamma\left(\sum_i \delta_{x_i,l_i}\right) = \frac{1}{\lambda}\sum_{l_i} \delta_{x_i}$$

The moment condition (1.11) is satisfied if we consider ψ of the form

$$\psi(x,l) = 1 + \sum_i \psi_B(x_i)(1+a+b)e^{-l_i}, \tag{4.16}$$

so that the averaged $\tilde{\psi}$ is of the form

$$\tilde{\psi}(x) = 1 + \sum_i \psi_B(x_i)(1+a+b)(1-e^{-\lambda}). \tag{4.17}$$

We note that the 1 appearing in the definitions of ψ and $\tilde{\psi}$ has been added only to ensure that both of these functions are greater than or equal to 1. □

We also investigate the behaviour of the lookdown process with generator A_λ as the parameter λ tends to infinity. A naive limiting procedure does not lead to a well defined object, since the form of the generator does not take into account the cancellations coming from fluctuations in the environment. In order to identify the correct limit, we once again use the separation of timescales trick.

Our argument is built around the Markov Mapping Theorem and Theorem 3.16. Our presentation is focused on the part of the proof which appeals to the Markov Mapping Theorem.

Remark 4.13. *Our computations do not lead to any surprising results - we show that the SBMRE can be obtained as a scaling limit of BBMRE. Even though we do not show convergence of the lookdown constructions, our argument, combined with tightness of the sequence of projected processes, guarantees convergence of the sequence of projected models. We therefore provide a new construction for the SBMRE. However, we still refer to results described in Remark 4.5 to guarantee the uniqueness of solutions to the projected martingale problem. The question of uniqueness of solutions to the martingale problem for the limiting process with levels (which, by the Markov Mapping Theorem would guarantee uniqueness of solutions to the projected model) will be pursued elsewhere.*

Proof[of Theorem 4.10] We define a test function of the form

$$h_1(\zeta,x,l) = -f_1(x,l)b\sum_i \zeta(x_i)l_i\frac{\partial_l g(x_i,l_i)}{g(x_i,l_i)}$$

and apply the generator (4.10) to a test function of the form $G = f_1 + \frac{1}{\sqrt{\lambda}}h_1$. This leads to

$$\begin{aligned}
A_\lambda\Big(f_1(x,l)+\frac{1}{\sqrt{\lambda}}h_1\Big) &= f_1(x,l)\sum_i \frac{Bg(x_i,l_i)}{g(x_i,l_i)}\\
&+ f_1(x,l)\sum_i 2a\int_{l_i}^{\lambda}(g(x_i,v)-1)\mathrm{d}v + f(x,l)\sum_i(al_i^2-\sqrt{\lambda}\zeta(x_i)bl_i)\frac{\partial_l g(x_i,l_i)}{g(x_i,l_i)}\\
&+\sqrt{\lambda}f_1(x,l)\sum_i\zeta(x_i)l_i\frac{\partial_l g(x_i,l_i)}{g(x_i,l_i)}\\
&-\frac{1}{\sqrt{\lambda}}\Bigg\{f_1(x,l)b\zeta(x_i)\sum_i\Big[\frac{Bg(x_i,l_i)}{g(x_i,l_i)}+\frac{B\partial_l g(x_i,l_i)}{g(x_i,l_i)}\Big]\\
&+\Big[\sum_i\zeta(x_i)l_i\frac{\partial_l g(x_i,l_i)}{g(x_i,l_i)}\Big]f_1(x,l)\sum_i 2a\int_{l_i}^{\lambda}(g(x_i,v)-1)\\
&+f_1(x,l)\sum_i[al_i^2-\sqrt{\lambda}b\zeta(x_i)l_i]\Bigg(\sum_{j\neq i}l_j\zeta(x_j)\frac{\partial_l g(x_i,l_i)\partial_l g(x_j,l_j)}{g(x_i,l_j)g(x_i,l_i)}\\
&+\frac{\zeta(x_i)\partial_l g(x_i,l_i)+\zeta(x_i)l_i\partial^2_{l^2}g(x_i,l_i)}{g(x_i,l_i)}\Bigg)\Bigg\},
\end{aligned}\tag{4.18}$$

where we have used the fact that f_1 does not depend on the environment and that $\mathbb{E}_\pi[h_1]=0$, (where $\mathbb{E}_\pi$ is the expected value over the stationary distribution for the environment) since $\mathbb{E}_\pi[\zeta]=0$. Passing to the limit in (4.18) as λ tends to infinity we obtain

$$\begin{aligned}
A_\infty(f_1) &= f_1(x,l)\sum_i\frac{Bg(x_i,l_i)}{g(x_i,l_i)}\\
&+f_1(x,l)\sum_i 2a\int_{l_i}^{\infty}(g(x_i,v)-1)\mathrm{d}v + f_1(x,l)\sum_i(al_i^2-b^2l_i)\frac{\partial_{l_i}g(x_i,l_i)}{g(x_i,l_i)}\\
&-f_1(x,l,n)\sum_i b^2\zeta(x_i)l_i\Bigg(\sum_{j\neq i}\zeta(x_j)l_j\frac{\partial_{l_i}g(x_i,l_i)\partial_{l_j}g(x_j,l_j)}{g(x_i,l_j)g(x_i,l_i)}+\frac{\zeta(x_i)l_i\partial^2_{l_i^2}g(x_i,l_i)}{g(x_i,l_i)}\Bigg).
\end{aligned}$$

Conditions of Theorem 3.16 are satisfied if we consider $A = A_\infty(f_1)$ and $\varepsilon_n^f = \mathcal{O}\Big(\frac{1}{\lambda}\Big)$. We now show that if we average the levels of the limiting generator, we obtain the

generator of the SBMRE of Definition 4.4. The general principle is the same as for the proof of Theorem 4.9 - we average out the levels and refer to the Markov Mapping Theorem to show that the distribution of the projected process is the distribution of the SBMRE.

We consider a Poisson random measure with distribution $\alpha(\mu, \mathrm{d}x \times \mathrm{d}l)$ on $\mathbb{R}^d \times \mathbb{R}_+$ with mean measure $\mu \times m_{leb}$, where m_{leb} is the Lebesgue measure. Just as in the Branching Brownian motion case, we consider a special set of test functions of the form

$$h(x) = \int_0^\infty (1 - g(x, l))\mathrm{d}l$$

In this setup, for a test function f, the projected (averaged) test function $\hat{f}$ takes form

$$\hat{f}(\mu) = \alpha f(\mu) = \int f(x, v)\alpha(\mu, \mathrm{d}x \times \mathrm{d}v) = e^{\int_{\mathbb{R}^d} \int_0^\infty (1-g(x,v))\mathrm{d}v\mu(\mathrm{d}x)} = e^{-\langle h, \mu\rangle},$$

which is a simple consequence of properties of Poisson random measures.

Once again, we integrate the groups of terms that behave similarly separately. Also, to make the calculations easier to read, we write the averaging 'level by level' - performing the computation for a single level wherever possible.

Since the part of the generator which describes the movement of the particles does not depend on the value of the level l the averaging is simply

$$\alpha\left(f_1(x, l) \sum_i \frac{Bg(x_i, l_i)}{g(x_i, l_i)}\right) = \int_{\mathbb{R}^d} -Bh(y)\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle}, \tag{4.19}$$

We now turn our attention to the terms which behave in a very similar fashion to those in (4.15). The computation is analogous.

$$\begin{aligned}
\int_{\mathbb{R}^d} \int_0^\infty avg(y, z) \int_v^\infty (1 - g(y, z))\mathrm{d}z\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
+ \int_{\mathbb{R}^d} \int_0^\infty (av^2 - b^2 v)\partial_v g(y, v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
= \int_{\mathbb{R}^d} \int_0^\infty avg(y, z) \int_v^\infty (1 - g(y, z))\mathrm{d}z\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
- \int_{\mathbb{R}^d} \int_0^\infty (2av - b^2)g(y, v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle}
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^d} \int_0^\infty avg(y,z) \int_v^\infty (1-g(y,z))\mathrm{d}z\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&- \int_{\mathbb{R}^d} \int_0^\infty 2a \int_v^\infty g(z,v)\mathrm{d}z\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} + \int_{\mathbb{R}^d} \int_0^\infty b^2 g(y,v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&= \int_{\mathbb{R}^d} a\bigg(\int_0^\infty [g(y,v)-1]\mathrm{d}v\bigg)^2 \mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} + \int_{\mathbb{R}^d} \int_0^\infty b^2 g(y,v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&= \int_{\mathbb{R}^d} \{ah^2(y) + b^2 h(y)\}\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle}, \quad (4.20)
\end{aligned}
$$

where we have integrated by parts and used the analogues of identities (4.13), (4.14).

Finally, the projections of the terms which are a direct consequence of separation of timescales lead to

$$
\begin{aligned}
&\int_{\mathbb{R}^d\times\mathbb{R}} b^2 \int_0^\infty \int_0^\infty \zeta(y_1)\zeta(y_2) v\partial_v g(y,v) z\partial_z g(y,z)\mathrm{d}v\mathrm{d}z\mu(\mathrm{d}y_1)\mu(\mathrm{d}y_2)e^{-\langle h,\mu\rangle} \\
&= -\int_{\mathbb{R}^d\times\mathbb{R}^d} b^2\zeta(y_1)\zeta(y_2) \int_0^\infty (g(y,v)-1)\mathrm{d}v \int_0^\infty (g(y,z)-1)\mathrm{d}v\mathrm{d}z\mu(\mathrm{d}y_1)\mu(\mathrm{d}y_2)e^{-\langle h,\mu\rangle} \\
&= \int_{\mathbb{R}^d\times\mathbb{R}^d} b^2\zeta(y_1)\zeta(y_2)h(y_1)h(y_2)\mu(\mathrm{d}y_1)\mu(\mathrm{d}y_2)e^{-\langle h,\mu\rangle}, \quad (4.21)
\end{aligned}
$$

where we have integrated by parts, and

$$
\begin{aligned}
&\int_{\mathbb{R}^d} b^2 \int_0^\infty v^2\partial_v^2 g(y,v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} = -\int_{\mathbb{R}^d} 2b^2 \int_0^\infty v\partial_v g(y,v)\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&= -\int_{\mathbb{R}^d} 2b^2 \int_0^\infty (1-g(y,v))\mathrm{d}v\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} = -\int_{\mathbb{R}^d} 2b^2 h(y)\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle}, \quad (4.22)
\end{aligned}
$$

where we have integrated by parts twice.

Combining the calculations (4.19), (4.20), (4.21), (4.22) and appealing to Theorem 3.16 to average over the environment we arrive at

$$
\begin{aligned}
\mathcal{L}\hat{f}(\mu) = \mathcal{L}e^{\langle f,\mu\rangle} &\int_{\mathbb{R}^d} \Big\{-Bh(y) + ah^2(y) - b^2 h(y)\Big\}\mu(\mathrm{d}y)e^{-\langle h,\mu\rangle} \\
&+ \int_{\mathbb{R}^d\times\mathbb{R}^d} b^2 q(y_1,y_2)h(y_1)h(y_2)\mu(\mathrm{d}y_1)\mu(\mathrm{d}y_2)e^{-\langle h,\mu\rangle},
\end{aligned}
$$

which is the generator of the SBMRE.

As before, to ensure that the solution of the martingale problem for the lookdown process gives us information about the solution of the martingale problem for the

projected process, we need to specify the Markov map γ and check that the conditions of the Markov Mapping Theorem are satisfied. Once again we appeal to Condition 4.11.

The Markov map γ is given by

$$\gamma\left(\sum_i \delta_{x_i,l_i}\right) = \begin{cases} \lim_{\lambda\to\infty} \frac{1}{\lambda} \sum_{l_i\leq\lambda} \delta_{x_i} & \text{if measures converge,} \\ \mu_0 & \text{otherwise .} \end{cases}$$

Our class of test functions is separating over the counting measures, and closed under multiplication. The moment condition 1.11 is satisfied if we consider ψ of the form (4.16) and the averaged $\tilde{\psi}$ of the form (4.17). □

4.3 Spatial Lambda-Fleming-Viot model

In this section we describe two different constructions of the SLFV model. The first one is a special case of the construction developed in Etheridge and Kurtz (2018), Section 4.1.3. This construction forms the basis for the construction of the SLFV with selection in a fluctuating environment, whose scaling limits we investigate in Section 4.5. The second one, which was the first lookdown construction for the SLFV, is based on the ideas of Véber and Wakolbinger (2015), and was developed using a different approach in Etheridge and Kurtz (2018), Section 4.1.1. Even though we do not use this approach in the later part of the thesis, it is still worth mentioning. The construction is much closer in flavour to the original ideas of Donnelly and Kurtz (1996). Moreover, inclusion of this example demonstrates the non-uniquness of lookdown representations - many different constructions may lead to the same averaged model. In both cases we restrict our attention to high density limits and models without selection. In both cases we do not intend to present any proofs or details.

We begin with some common elements of both constructions. Let us recall the key elements of the SLFV process. We consider a population living in a geographical space, which, for simplicity, we choose to be $\mathbb{R}^d$. Each individual is assigned a type from a typespace $\mathcal{K}$. Let $\mu = \Lambda \times \nu^1(w, \mathrm{d}u) \times \nu^2(\mathrm{d}w)$ be a measure on $\mathbb{R}^d \times [0,1] \times [0,\infty)$, where Λ is d-dimensional Lebesgue measure, ν^1 is a measure which determines *impacts*

of the events and ν^2 is a σ-finite measure of event radii which satisfy conditions which are specified in (4.23). Evolution of the population is driven by a Poisson point process Π on $[0,\infty)\times\mathbb{R}^d\times[0,1]\times[0,\infty)$ with mean measure $\Lambda\times\mu$. Whenever $(t,x,u,r)\in\Pi$, a reproduction event occurs at time t in the closed ball $B_r(x)$ (a ball of radius r centered at x) with *impact* u. The impact of the event determines the proportion of the individuals within the ball $B_r(x)$ that are replaced during the event by the offspring of a parent chosen from the ball $B_r(x)$ just before the event. The locations of new individuals are distributed uniformly over $B_r(x)$. For the construction to be valid, we assume that

$$\int_{[0,1]\times(0,\infty)} uw^d\nu^1(w,\mathrm{d}u)\nu^2(\mathrm{d}w)<\infty. \tag{4.23}$$

For simplicity, we only describe the construction for a fixed impact u and assume that the radius of the reproduction events is always fixed and equal to r.

In the spirit of lookdown constructions, in addition to a location in geographical space and a type in the typespace $\mathcal{K}$, each individual is equipped with a level $l\in\mathbb{R}_+\cup\{0\}$. The value of the level impacts the choice of the parent during a reproduction event.

As was the case for the models of Section 4.2, it is convenient to consider our model as a counting measure on $\mathbb{R}^d\times\mathcal{K}\times[0,\infty)$, where the first component encodes the geographical space, the second encodes the type of the individual and the third encodes the level of the individual. The state of the population is given by

$$\eta=\sum_i\delta_{x_i,\kappa_i,l_i},$$

and the single individual i is described by a triple (x_i,κ_i,l_i), where x_i is the location of the individual, κ_i is their type and l_i is their level. In our particular case the levels will always be a conditionally Poisson system with Cox measure $\Xi\times\Lambda$, where Λ is the Lebesgue measure on $\mathbb{R}$. The measure Ξ is then nothing else but the distribution of locations and types of individuals.

We specify the models in terms of generators. Let us describe the domains on which our generators are defined, which turns out to be useful not only for formalizing the constructions in this section but also will serve as a functional setup for the considerations

in Section 4.5. Define

$$\mathcal{D}_\lambda = \Bigg\{ f(\eta) = \prod_{x,\kappa,l\in\eta} g(x,\kappa,l) : \\ 0 \le g(x,\kappa,l) \le 1, g(\cdot,\kappa,l) \in C^2(\mathbb{R}^d), \|\partial_l g(x,\kappa,l)\| < \infty \\ \exists \text{ compact } K_g \in \mathbb{R}^d, 0 < l_g \le \lambda \\ g(x,\kappa,l) = 1 \text{ for } (x,l) \notin K_g \times [0,l_g] \Bigg\}, \quad (4.24)$$

and

$$\mathcal{D}_\infty = \bigcup_\lambda \mathcal{D}_\lambda. \quad (4.25)$$

Our test functions are specified by (4.25).

Remark 4.14. *Notice that our restrictions on domains of the generators are very similar to those in Condition 4.11. This is due to the fact that once again we will apply the Markov Mapping Theorem.*

We now describe the basic principles of the construction from Etheridge and Kurtz (2018), Section 4.1.2. The evolution of the population is based on events which are composed of two elements - discrete births and so-called thinning of the population. Whenever $(t,x) \in \Pi$, if the number of individuals within $B_r(x)$ is greater than zero, the birth event produces offspring, with levels distributed on $[0,\infty)$ according to independent Poisson point processes with intensity $\alpha_z = uV_r$ (recall that V_r denotes the volume of ball of radius r). The levels of new particles are denote by $(v_1, v_2, \dots, v_k, \dots)$. Their locations are distributed uniformly over $B_r(x)$. Let v^* be the minimum of $(v_1, v_2, \dots, v_k, \dots)$.

Let (x^*, κ^*, l^*) denote the element in η such that $x^* \in B_r(y)$ with the smallest level greater than v^*. The individual (x^*, κ^*, l^*) is chosen as the parent of the event and removed from the population. All new individuals are assigned a type which is same as the type of the parent. The levels of old individuals in the population are changed. If the level of the individual was smaller than v^*, it remains unaffected by the birth part of the event. If the level of the individual was larger than v^*, it is moved to $l - l^* + v^*$). The thinning occurs after the movement of the levels due to birth event has been accounted

for. The new individuals are not affected by thinning. The thinning takes the new level of each individual present within the ball $B_r(x)$ just before the event (apart from the parent), and multiplies it by $1/(1-u)$.

We note that instead we can identify the parent with the lowest offspring (with level v^*) rather than removing it from the population. The choice between those two options is a matter of convenience and does not affect the model. In our considerations in Section 4.4 and Section 4.5 it is more convenient to not remove the parent.

Let $v_{y,r}$ denote the density of the uniform distribution on $B_r(y)$. Let $\mathcal{J}_{EK}$ denote the expected value of the test function evaluated immediately after an event centred at y. It is given by

$$\begin{aligned}\mathcal{J}_{EK}(g,\eta) = \prod_{(x,\kappa,l)\in\eta, x\notin B_r(y)} g(x,\kappa,l) \\ \times \int_0^\infty \Bigg[\alpha_z e^{-\alpha_z v^*} \int g(x',\kappa^*,v^*) v_{y,r}(\mathrm{d}x') \\ \times \exp\Bigg(-\alpha_z \int_{v^*}^\infty \Bigg(1 - \int g(x',\kappa^*,v^*) v_{y,r}(\mathrm{d}x')\Bigg)\mathrm{d}v^*\Bigg) \\ \times \prod_{(x,\kappa,l)\in\eta, x\in B_r(y), l>l^*} g(x,\kappa,\frac{1}{1-u}(l-l^*+v^*)) \\ \times \prod_{(x,\kappa,l)\in\eta, x\in B_r(y), l<l^*} g\Bigg(x,\kappa,\frac{1}{1-u}l\Bigg)\Bigg]\mathrm{d}v^*\end{aligned}$$

The generator of the lookdown representation of the SLFV can be now written as

$$A_{EK}f(\eta) = \int_{\mathbb{R}^d} \mathbf{1}_{\eta(B_r(y)\times[0,\infty))>0}\{\mathcal{J}_{EK}(g,\eta) - f(\eta)\}\mathrm{d}y. \tag{4.26}$$

Recall that η is a conditionally Poisson process with Cox measure $(\Xi(s)\times\Lambda)$. To average the generator over the distribution of the levels, we define

$$h(x,\kappa) = \int_0^\infty (1-g(x,\kappa,l))\mathrm{d}l \tag{4.27}$$

and

$$h^*_{y,r}(\kappa) = \int \Bigg(1 - \int g(x',\kappa,l) v_{y,r}(\mathrm{d}x')\Bigg)\mathrm{d}l. \tag{4.28}$$

Observe that, by integration by parts,

$$\int_0^\infty \Big\{ \alpha_z e^{-\alpha_z v^*} \int g(x', \kappa^*, v^*) v_{y,r}(\mathrm{d}x') \\ \times \exp\Big(-\alpha_z \int_{v^*}^\infty \Big(1 - \int g(x', \kappa^*, v^*) v_{y,r}(\mathrm{d}x') \Big) \mathrm{d}v \Big) \Big\} \mathrm{d}v^* = e^{-\alpha_z h^*_{y,r}(\kappa^*)}$$

Therefore if we average out the levels in generator (4.26) we obtain

$$\alpha A_{EK} f(\Xi) = \exp\Big(\int_{\mathbb{R}^d} h(x,\kappa) \Xi(\mathrm{d}x, \mathrm{d}\kappa) \Big) \\ \times \int_{\mathrm{R}^d} \Big\{ \mathbb{H}_2(h^*_{y,r}, \Xi) \exp\Big(u \int_{B_r(x)\times\mathcal{K}} h(x,\kappa) \Xi(\mathrm{d}x, \mathrm{d}\kappa) \Big) - 1 \Big\} \mathrm{d}y,$$

where

$$\mathbb{H}_2(h^*_{y,r}, \Xi) = \frac{1}{\Xi(B_r(x)\times\mathcal{K})} \int_{B_r(x)\times\mathcal{K}} \exp\Big(-u V_r h^*_{y,r}(\kappa) \Big) \Xi(\mathrm{d}x, \mathrm{d}\kappa).$$

The averaged model is just the usual Spatial Lambda-Fleming-Viot model, see Etheridge and Kurtz (2018), Section 4.1. We also observe that if $\Xi(0, \mathrm{d}x \times \mathcal{K})$ is the Lebesgue measure, so is $\Xi(t, \mathrm{d}x \times \mathcal{K})$, for arbitrary t. If this is the case, $\mathbb{H}_2$ can be written as

$$\mathbb{H}_2(h^*_{y,r}) = \frac{1}{\Xi(B_r(x)\times\mathcal{K})} \int_{B_r(x)\times\mathcal{K}} \exp\Big(-u h^*_{y,r}(\kappa) \Xi(B_r(x)\times\mathcal{K})) \Big) \Xi(\mathrm{d}x, \mathrm{d}\kappa).$$

A different lookdown construction for the SLFV has been proposed by Véber and Wakolbinger (2015).

The Poisson point process Π driving the evolution is the same as before. However, what happens during reproduction events, and the evolution of the level processes, are very different.

Whenever a reproduction event occurs, each individual within the ball $B_r(x)$ participates in the event with probability u. All individuals that participate in the event are moved to new positions within the ball $B_r(x)$, with their spatial locations sampled (independently) from the uniform distribution over $B_r(x)$. Their type is set to be the same as the type of the individual with the lowest level among those participating in the event.

Before we proceed to specifying the generator of the process, we need to introduce some more notation. Let $\eta_{y,r}$ denote the measure corresponding to all individuals with locations within the ball $B_r(y)$, that is

$$\eta_{y,r} = \sum_{(x,\kappa,l):x\in B_y(r)} \delta_{x,\kappa,l}.$$

Let S denote the set of all individuals involved in the event and $\kappa^*(S)$ be the type of the individual with the lowest level among them. Let $\mathcal{J}_{VW}(g,\widehat{g},S,y)$ denote the expected value of the test function evaluated immediately after an event centred at y. It is given by

$$\mathcal{J}_{VW}(g,S,y) = \prod_{(x,\kappa,l)\in S} u\Big(\int_g (x',\kappa^*(S),l)v_{y,r}(\mathrm{d}x')\Big) \prod_{(x,\kappa,l)\notin S\cap(x,\kappa,l)\in\eta_{y,r}} \{(1-u)g(x,\kappa,l)\}.$$

The generator of this process is given by

$$A_{VW}f(\eta) = f(\eta)\int_{\mathbb{R}^d\times[0,1]\times[0,\infty)} \left(\frac{\sum_{S\in\eta_{y,r}} \mathcal{J}_{VW}(g,\widehat{g},S,y)}{\prod_{(x,\kappa,l)\in\eta_{y,r}} g(x,\kappa,l)} - 1\right)\mathrm{d}y.$$

Let h and h^* be defined as in (4.27) and (4.28), respectively. The averaged version of this model is

$$\alpha A_{EK}f(\Xi) = \exp\Big(\int_{\mathbb{R}^d} h(x,\kappa)\Xi(\mathrm{d}x,\mathrm{d}\kappa)\Big) \times \int_{\mathrm{R}^d}\left\{\mathbb{H}_1(h^*_{y,r},\Xi)\exp\left(u\int_{B_r(x)\times\mathcal{K}} h(x,\kappa)\Xi(\mathrm{d}x,\mathrm{d}\kappa)\right) - 1\right\}\mathrm{d}y,$$

with H_1 given by

$$\mathbb{H}_1(h^*_{y,r},\Xi) = \frac{1}{\Xi(B_r(x)\times\mathcal{K})}\int_{B_r(x)\times\mathcal{K}} \exp\Big(-uV_r h^*_{y,r}(\kappa)\Xi(B_r(x)\times\mathcal{K}))\Big)\Xi(\mathrm{d}x,\mathrm{d}\kappa).$$

We therefore observe that even though the two lookdown representations described in this subsection differ significantly, their averaged versions are the same.

4.4 Scaling limits of the LFV - dynamics of the rare type

In this section we are interested in describing the evolution of a subpopulation with a rare mutation within a population which evolves according to the Λ-Fleming-Viot model with selection in a fluctuating environment (LFVSFE). We therefore consider the population with two genetic types - the 'rare' type and the 'common' type (often refereed to as a wild type in biological literature). Which of those two types has higher fitness changes with the changes in the environment. The selection mechanism considered here differs slightly from the one of Section 3.1, nevertheless it is closely related , see Remark 4.23. We show that the evolution of the rare subpopulation follows a Feller diffusion in random environment, which is the non-spatial counterpart of the SuperBrownian motion in a random environment discussed in Section 4.2.

We provide a new description of the LFVSFE in terms of a lookdown construction. This construction is inspired by the lookdown construction of the neutral model in Subsection 4.1.3 of Etheridge and Kurtz (2018).

As is the case for the models from Section 4.3, each of the individuals in the population is assigned a *genetic type* κ from the set $\mathcal{K}$ and a *level* $l \in \mathbb{R}_+ \cup \{0\}$. We restrict our attention to $\mathcal{K} = \{\kappa_c, \kappa_r\}$, which we refer to as the 'common' and 'rare' type, respectively. Since in this section we consider a model without spatial structure, the state of the population can be represented as a collection of points, $\eta = \{(l, \kappa)\}$, or as a measure which can be written, with an abuse of notation, as

$$\eta = \sum_{(l,\kappa)\in\eta} \delta_{(l,\kappa)}, \quad \text{where } (l, \kappa) \in (\mathbb{R}_+ \cup \{0\}, \mathcal{K}).$$

Once again, we assume that the process with levels η is always a conditionally Poisson system with Cox measure $\Lambda \times \Xi$.This is because, in our lookdown model, when our initial condition has this form then our process will have this form for all subsequent times.

We are interested in the situation when which type is selectively advantageous depends on the environment. We write ζ for the random variable which models the

state of the environment. Therefore the full state of the model at time t is given by a pair (η_t, ζ_t).

The evolution of the population is determined by reproduction events of two types - neutral and selective. We assume that a proportion, s, of the events is selective and favours one of the two types. Events are driven by independent Poisson processes Π^{neu} and Π^{sel}. For simplicity we assume that the impact of the events is fixed and equal to u.

Both neutral and selective events are composed of two elements - discrete births and thinning. The birth phase of the events differs between neutral and selective events, while the thinning phase is the same.

Whenever $t \in \Pi^{neu}$, a birth event produces offspring, with levels distributed according to an independent Poisson point process with intensity u. Let v^* be the smallest of the new levels. Let (l^*_{neu}, κ^*) denote the element of η with the smallest level greater than v^*, that is

$$l^*_{neu} = \min\{l : (l, \kappa) \in \eta, l > v^*\}.$$

The individual (l^*_{neu}, κ^*) is chosen as the parent of the event and removed from the population. All new individuals are assigned a type which is the same as the type of the parent. The levels of all old individuals in the population are changed. If the level of the individual was smaller than v^* it remains unaffected by the birth. If the level of the individual was larger than v^*, it is moved to $l - l^*_{neu} + v^*$.

If the event is selective, the situation is more complicated. We introduce an additional function $\sigma(\kappa, \zeta)$, which influences the likelihood of an individual of type κ being parent, given the state of the environment, ζ. When the environment is in state ζ, the higher the value of $\sigma(\kappa, \zeta)$, the more likely an individual of type κ is to become a parent.

Whenever $t \in \Pi^{sel}$, a birth event produces offspring, with levels distributed according to an independent Poisson process with intensity u. As before, let v^* be the smallest of the new levels. Let (l^*_{sel}, κ^*) denote the element of η which obtains the minimum

$$\min\left\{\frac{l_i - v^*}{\sigma(\kappa_i, \zeta)} : (l_i, \kappa_i) \in \eta, l_i > v^*\right\}. \tag{4.29}$$

The individual (l^*_{sel}, κ^*) is chosen as the parent of the event and removed from the population. All new individuals are assigned a type which is the same as the type of the parent. The levels of all old individuals in the population are changed. If the level of the individual was smaller than v^* it remains unaffected by the birth. If the level of the individual was larger than v^*, it is moved to $\sigma(\kappa, \zeta)(l - l^*_{neu} + v^*)/\sigma(\kappa^*, \zeta)$.

For both neutral and selective events once the parent has been selected, and the old levels moved to their new locations, thinning takes place. Thinning does not affect the new individuals. The thinning takes the new level of each individual present within the population just before the event (excluding the parent) and multiplies it by $1/(1-u)$. The combined effect of this movement is defined explicitly through the function $\mathcal{J}_{neu}$, defined in (4.31), for neutral events and $\mathcal{J}_{sel}$, defined in (4.33), for selective events.

We are now ready to define the process of interest.

Definition 4.15 (Lookdown representation of LFVSFE)**.** *Fix $s \in (0,1)$. Let Π^{neu}, Π^{sel} be a pair of independent Poisson point processes with intensity measures $(1-s)\mathrm{d}t \otimes \nu(\mathrm{d}u)$ and $s\mathrm{d}t \otimes \nu(\mathrm{d}u)$ respectively on $\mathbb{R}^+ \times (0,1)$. Moreover, let Π^{env} be a Poisson process with rate E, independent of Π^{neu}, Π^{sel}. Let $\sigma : \mathcal{K} \times \{-1,1\} \to \mathbb{R}$ be a function.*

The lookdown representation of LFVSRE is the process taking values in purely atomic measures on $\mathbb{R} \times \mathcal{K} \times \{-1,1\}$ with dynamics described as follows.

1. *If $(t,u) \in \Pi^{neu}$*

   (a) *a group of new individuals with levels $(v_1, v_2, \dots)$ is added to the population. Their levels are distributed according to a Poisson process with intensity u.*

   (b) *Let $v^* = \min\{v_1, v_2, \dots\}$. The type of the new individuals is chosen to be the same as the type of the individual (κ^*, l^*) whose level is the lowest above v^*, that is*

$$l^*_{neu} = \min\{l : (l, \kappa) \in \eta, l > v^*\}. \tag{4.30}$$

(c) As a result of an event the levels with position l before an event will have new position given by

$$\mathcal{J}_{neu}(l, l^*_{neu}, v^*) = \begin{cases} \frac{1}{1-u}(l - (l^*_{neu} - v^*)) & \text{if } l > l^*_{neu}, \\ \frac{1}{1-u} l & \text{if } l < l^*_{neu}, \\ v^* & \text{if } l = l^*_{neu}. \end{cases} \tag{4.31}$$

2. If $(t, u) \in \Pi^{sel}$

(a) a group of new individuals with levels $(v_1, v_2, \dots)$ is added to the population. Their levels are distributed according to a Poisson process with intensity u.

(b) Let $v^* = \min\{v_1, v_2, \dots\}$. The type of the new individuals is chosen to be the same as the type of the individual (l^*, κ^*) whose level minimizes

$$\left\{ \frac{l_i - v^*}{\sigma(\kappa_i, \zeta)} : (l_i, \kappa_i) \in \eta, l_i > v^* \right\}. \tag{4.32}$$

(c) As a result of an event the levels with position l and type κ before an event will maintain their type but will have new position given by

$$\mathcal{J}_{sel}((l, \kappa), (l^*_{sel}, \kappa^*), \zeta, v^*) = \begin{cases} v^* & \text{if } l = l^*_{sel}, \\ \frac{1}{1-u}\Big(l - (l^* - v^*)\frac{\sigma(\kappa,\zeta)}{\sigma(\kappa^*,\zeta)}\Big) & \text{if } l \neq l^*_{sel}, l > v^*, \\ \frac{1}{1-u} l & \text{if } l < v^*. \end{cases} \tag{4.33}$$

3. If $t \in \Pi^{env}$, the environmental variable ζ_t is resampled uniformly from $\{-1, 1\}$.

Our results require a symmetry condition on the values of $\sigma(\kappa, \zeta)$. To be more precise, we require that

$$\mathbb{E}_\pi\left[\frac{\sigma(\kappa_r, \zeta)}{\sigma(\kappa_c, \zeta)} - 1\right] = 0, \tag{4.34}$$

where π is the stationary distribution of ζ.

We note that the generator of the process from Definition 4.15 is given by

$$Af(\eta,\zeta) = A_{neu}f(\eta,\zeta) + A_{sel}f(\eta,\zeta) + A_{env}f(\eta,\zeta), \tag{4.35}$$

where

$$\begin{aligned}
A_{neu}f(\eta,\zeta) =&(1-s)\Bigg(\int_0^\infty \Bigg[ue^{-uv^*}g(\kappa^*,v^*)e^{-u\int_{v^*}^\infty(1-g(\kappa^*,v))\mathrm{d}v} \\
&\times \prod_{(\kappa,l)\in\eta,l\neq l^*} g(\kappa,\mathcal{J}_{neu}(l,l^*_{neu},v^*))\Bigg]\mathrm{d}v^* - f(\eta)\Bigg), \\
A_{sel}f(\eta,\zeta) =&s\Bigg(\int_0^\infty \Bigg[ue^{-uv^*}g(\kappa^*,v^*)e^{-u\int_{v^*}^\infty(1-g(\kappa^*,v))\mathrm{d}v} \\
&\times \prod_{(\kappa,l)\in\eta,l\neq l^*_{sel}} g(\kappa,\mathcal{J}_{sel}((l,\kappa),(l^*_{sel},\kappa^*),\zeta,v^*))\Bigg]\mathrm{d}v^* - f(\eta)\Bigg), \\
A_{env}f(\eta,\zeta) =&\mathbb{E}_\pi[f(\eta,\zeta)] - f(\eta,\zeta).
\end{aligned}$$

Observe that the only difference between neutral and selective events is in the choice of the parent l^* and the movement of levels.

4.4.1 Scaling

We are interested in the evolution of the subpopulation of the rare type within the population evolving according to LFVSFE. To quantify the rarity, we assume that we consider a population with total density K (which is equal to one for the models of Chapter 3), and will let K tend to infinity. We wish the rare type to make up about $\mathcal{O}(1/K)$ of the population at each level of the scaling. This is represented in the look-down process by the intensity of levels. The levels of individuals are initially Poisson distributed with intensity K with $\mathcal{O}(1)$ individuals of the rare type and the rest of the common type.

In order to recover the correct scaling limit we need to readjust our parameters. Let w^N denote the proportion of individuals of the rare type at the N^{th} stage of scaling. Let $X^N = Kw^N$ denote the size of the population of the rare type. Our scaling limit describes the behaviour of X^N. The right scaling of the other parameters is suggest by Chetwynd-Diggle and Etheridge (2018). We speed up the reproduction rate by N and

increase the total population size to $K(N)$, but scale down both the impact and the selection coefficient. The impact of an event at the N^{th} stage of the approximation will be given by $u/J(N)$. The selection coefficient in the presence of fluctuations will be $\widehat{S}(N)s/S(N)$. When we later prove Theorem 4.21 we will consider the case without a fluctuating environment. In this case, the selection coefficient is given by $s/S(N)$. In order to simplify the notation we drop the explicit dependence of scaling parameters on N in what follows. For our results to hold, certain relations between the parameters need to be satisfied. In Theorem 4.18, Theorem 4.20 and Theorem 4.21 we specify scaling limits for the model with fluctuating selection, the model where the direction of selection does not change and the neutral model respectively. In particular, as was the case in Section 3.1, our results require a specific relation between the rate of the changes in the environment and the rate of selection. We therefore assume that the rate of the environmental events is $\widehat{S}^2$, which is consistent with the scaling in Chapter 3, see discussion above (3.6).

The generator of the scaled process can be written as

$$A^N f(\eta,\zeta) = A^N_{neu} f(\eta,\zeta) + \widehat{S} A^N_{sel} f(\eta,\zeta) + \widehat{S}^2 A^N_{env} f(\eta,\zeta), \tag{4.36}$$

where

$$\begin{aligned} A^N_{neu} f(\eta) = N\Bigg(\int_0^\infty \Bigg[&\frac{uK}{J} e^{-\frac{uK}{J}v^*} g(\kappa^*, v^*) e^{-\frac{uK}{J}\int_{v^*}^\infty (1-g(\kappa^*,v))\mathrm{d}v} \\ &\times \prod_{(\kappa,l)\in\eta, l\neq l^*} g(\kappa, \beth_{neu}(l,l^*,v^*)) \Bigg] \mathrm{d}v^* - f(\eta) \Bigg), \\ A^N_{sel} f(\eta) = \frac{sN}{S}\Bigg(\int_0^\infty \Bigg[&\frac{uK}{J} e^{-\frac{uK}{J}v^*} g(\kappa^*, v^*) e^{-\frac{uK}{J}\int_{v^*}^\infty (1-g(\kappa^*,v))\mathrm{d}v} \\ &\times \prod_{(\kappa,l)\in\eta, l\neq l^*_{sel}} g(\kappa, \beth_{sel}(l,l^*_{sel},v^*)) \Bigg] \mathrm{d}v^* - f(\eta) \Bigg). \end{aligned}$$

4.4.2 Statement of main results

In this subsection we state the main results of this section. In order to state the results formally, we recall the definitions of the non-spatial counterparts of the models from Section 4.2. We specify them in terms of their generators. For their lookdown

representations, it is convenient to consider test functions of a form analogous to (4.3), that is

$$f(l,n) = \prod_{i=1}^{n} g(l_i),$$

with the additional requirement that there exists a $\lambda_g > 0$ such that $g(l) = 1$ for each $l > \lambda_g$. This set of test functions will be used for all computations involving lookdown representations in this section.

Definition 4.16 (Feller diffusion)**.** *Let $a, b > 0$. The Feller diffusion is the process taking values in $\mathbb{R}$ with generator C_{fd}, defined for every $f \in C^\infty(\mathbb{R})$, given by*

$$C_{fd}f(y) = ayf''(y) + byf'(y).$$

Its lookdown representation is characterised by the process with generator A_{fd} given by

$$A_{fd}f(\eta) = f(\eta)\sum_i 2a\int_{l_i}^{\infty}(g(v)-1)\mathrm{d}v + f(\eta)\sum_i (al_i^2 - bl_i)\frac{g'(l_i)}{g(l_i)}.$$

Definition 4.17 (Feller diffusion in random environment)**.** *Let $a, b > 0$. The Feller diffusion in a random environment is the process taking values in $\mathbb{R}$ with generator C_{fdr}, defined for every $f \in C^\infty(\mathbb{R})$, of the form*

$$C_{fdr}f(y) = (ay + b^2y^2)f''(y) + b^2yf'(y).$$

Its lookdown representation is characterised by the process with generator A given by

$$\begin{aligned} Af(\eta) = f(\eta)\sum_i 2a\int_{l_i}^{\infty}(g(v)-1)\mathrm{d}v + f(\eta)\sum_i (al_i^2 - b^2l_i)\frac{g'(l_i)}{g(l_i)} \\ + b^2 f(\eta)\sum_j\left(\sum_{i\neq j} l_jl_i\frac{g'(l_i)g'(l_j)}{g(l_i)g(l_j)} + l_j^2\frac{g''(l_i)}{g(l_i)}\right). \end{aligned} \quad (4.37)$$

For a detailed discussion of the lookdown constructions for the Feller diffusion and the Feller diffusion in a random environment we refer to Kurtz and Rodrigues (2011), Section 2.

Theorem 4.18. *Let $X^N(t)$ denote the total intensity of individuals of rare type at time t. Suppose that X_0^N converges to X_0. Moreover, suppose that, as N tends to infinity,*

$$J, K, S, \widehat{S} \to \infty; \quad \frac{K}{J} \to 0; \quad \frac{u^2 NK}{J^2} \to 2a; \quad \frac{N^2}{KJ^2} \to 0;$$
$$\mathbb{E}_\pi\left[\left\{\frac{suN}{SJ}\left(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)} - 1\right)\right\}^2\right] \to b^2; \quad \frac{\widehat{S}}{K} \to 0; \quad \frac{\widehat{S}}{S} \to 0. \tag{4.38}$$

In addition, assume that there exists an m such that, as $N \to \infty$, $NK^m/J^m \to 0$. Then the sequence $X_N(t)$ converges weakly to the Feller diffusion in a random environment with initial condition X_0 and parameters $2a$, b^2.

Example 4.19. *Fix $\epsilon \in (0, 1/4)$, $\beta \in (0, 1/4 - \epsilon)$ and $\gamma \in (0, \beta)$. The conditions of Theorem 4.18 are satisfied if*

$$J = N^{\frac{3}{4}+\epsilon} \quad S = N^\beta \quad K = N^{\frac{1}{2}+2\epsilon} \quad \widehat{S} = N^\gamma.$$

As a by-product of our technique, we prove an analogous result for the neutral model and the model with selection.

Theorem 4.20. *Let $X^N(t)$ denote the total intensity of individuals of rare type at time t. Suppose that X_0^N converges to X_0 and the intensity of selective events is zero. Moreover, suppose that, as N tends to infinity,*

$$J, K \to \infty; \quad \frac{K}{J} \to 0; \quad \frac{N^2}{KJ^2} \to 0; \quad \frac{u^2 NK}{J^2} \to 2a. \tag{4.39}$$

In addition, assume that there exists an m such that, as $N \to \infty$, $NK^m/J^m \to 0$. Then the sequence $X_N(t)$ converges weakly to the critical Feller diffusion ($b = 0$) with initial condition X_0 and variance parameter $2a$.

Theorem 4.21. *Let $X^N(t)$ denote the total intensity of individuals of rare type at time t. Suppose that X_0^N converges to X_0 and σ does not depend on the environment. Moreover, suppose that, as N tends to infinity,*

$$J, K, S \to \infty; \quad \frac{K}{J} \to 0; \quad \frac{N^2}{KJ^2} \to 0; \quad \frac{u^2 NK}{J^2} \to 2a; \quad \frac{suN}{SJ}\left(\frac{\sigma(\kappa_r)}{\sigma(\kappa_c)} - 1\right) \to b \tag{4.40}$$

In addition, assume that there exists an m such that, as $N \to \infty$, $NK^m/J^m \to 0$. Then the sequence $X_N(t)$ converges weakly to the Feller diffusion with initial condition X_0 and parameters $2a$, b.

Remark 4.22. *We stress that our proofs do not guarantee convergence of the lookdown representations, but only convergence of the projected models.*

4.4.2.1 Projected model

We follow the work in Etheridge and Kurtz (2018) to show the connection between our lookdown construction and the standard LFV.

The neutral generator is simply the non-spatial counterpart of that in Etheridge and Kurtz (2018) and so we do not treat it here. However our selection is not one of the standard forms of selection and so we investigate further. This section follows Etheridge and Kurtz (2018) and all the techniques in this section are taken from there, but formulated in our notation for completeness. The part of the generator of our process which describes selective events is of the form

$$A_{sel}f(\eta) = s\Bigg(\int_0^\infty \Bigg[u e^{-uv^*} g(v^*, \kappa^*) e^{-u\int_{v^*}^\infty (1-g(v,\kappa^*))\mathrm{d}v} \times \prod_{(l,\kappa)\in\eta, l\neq l^*_{sel}} g(\mathcal{J}_{sel}(l, l^*_{sel}, v^*), \kappa) \Bigg] \mathrm{d}v^* - f(\eta) \Bigg).$$

We define $h(\kappa) = \int_0^\infty (1 - g(l,\kappa))\mathrm{d}l$ and note that integration by parts gives

$$\int_0^\infty u e^{-uv^*} g(v^*, \kappa^*) e^{-\frac{uK}{J}\int_{v^*}^\infty (1-g(v,\kappa^*))\mathrm{d}v} \mathrm{d}v^* = e^{-uh(\kappa^*)}.$$

We also note that if η is formed from a Poisson point process with intensity measure $\Xi(\mathrm{d}\kappa)$ then

$$\{\mathcal{J}_{sel}(l, l^*_{sel}, v^*) : l \neq l^*_{sel}\},$$

is a Poisson point process with intensity measure $(1-u)\Xi(\mathrm{d}\kappa)$. Consider the distribution of κ^* conditioned on $\Xi(\mathrm{d}\kappa)$. We note that

$$\mathbb{P}[\kappa^* = \kappa_r] = \frac{\sigma(\kappa_r)\Xi(\{\kappa_c\})}{\sigma(\kappa_r)\Xi(\{\kappa_c\}) + \sigma(\kappa_c)\Xi(\{\kappa_r\})},$$

and so when we average our selective generator we get

$$\begin{aligned}\alpha A^N_{sel} f(\Xi) = s e^{-\int_{\mathfrak{K}} h(\kappa)\Xi(\mathrm{d}\kappa)} \\ \times \Bigg(\Bigg[\frac{\sigma(\kappa_r)\Xi(\{\kappa_c\})}{\sigma(\kappa_r)\Xi(\{\kappa_c\}) + \sigma(\kappa_c)\Xi(\{\kappa_r\})} e^{-uh(\kappa_r)} \\ + \frac{\sigma(\kappa_c)\Xi(\{\kappa_r\})}{\sigma(\kappa_r)\Xi(\{\kappa_c\}) + \sigma(\kappa_c)\Xi(\{\kappa_r\})} e^{-uh(\kappa_c)} \Bigg] e^{u\int_{\mathfrak{K}} h(\kappa)\Xi(\mathrm{d}\kappa)} - 1 \Bigg).\end{aligned}$$

We finally note that our selection can now be seen as a simple weighted choice of parent in the (non-spatial) Lambda Flemming-Viot process. The same calculation can be done for the spatial case.

Remark 4.23. *We notice that our way of modelling selection differs from the approach of Etheridge et al. (2014). However, a quick generator calculation shows that this type of selection leads to the same diffusion approximation as in Etheridge et al. (2014) and Chapter 3.*

4.4.3 Convergence of the non-spatial model

We are interested in convergence of the model with selection in a fluctuating environment. Recall that we study the behaviour of an establishing mutation under this model. We take that into account by considering test functions which are unaffected by the individuals of the common type, κ_c.

We recall the generator of the rescaled process takes the form

$$A^N f(\eta,\zeta) = A^N_{neu} f(\eta,\zeta) + \widehat{S} A^N_{sel} f(\eta,\zeta) + \widehat{S}^2 A^N_{env} f(\eta,\zeta), \tag{4.41}$$

where A^N_{neu} is the part of the generator which describes neutral events, A^N_{sel} is the part of the generator which describes selective events, and A^N_{env} describes the evolution of the environment.

In Subsections 4.4.3.2 and 4.4.3.3, we show that the terms A^N_{neu} and A^N_{sel} converge to well-defined limits. However, since $\widehat{S} \to \infty$ as N tends to infinity, it may seem that as N tends to infinity, (4.41) will not converge to a non-trivial limit. However, the naive limiting procedure does not take into account the changes in the direction of selection. In order to identify the correct limit, we apply a 'separation of timescales' trick due to Kurtz (1973).

By the calculations in the proof of Theorem 4.21 the operator A^N_{sel} can be written as

$$A^N_{sel} f(\eta) = f(\eta) \frac{suN}{JS} \left(\frac{\sigma(\kappa_r, \zeta)}{\sigma(\kappa_c, \zeta)} - 1 \right) \sum_j l_j \frac{g'(\kappa_j, l_j, \xi)}{g(\kappa_j, l_j, \xi)} + \mathcal{O}\left(\frac{1}{S} + \frac{1}{K} \right) \tag{4.42}$$

We consider a test function $\widetilde{f}$ of the form

$$\widetilde{f}(\eta, \zeta) = f(\eta) + \frac{1}{\hat{S}} f(\eta) \frac{suN}{JS} \left(\frac{\sigma(\kappa_r, \zeta)}{\sigma(\kappa_c, \zeta)} - 1 \right) \sum_j \frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)} l_j =: f(\eta) + \frac{1}{\hat{S}} f_1(\eta, \zeta).$$

Observe that since $\hat{S} \to \infty$, we can prove the test function $\widehat{f}$ will tend to $f(\eta)$ as $N \to \infty$.We apply the generator (4.41) to $\widehat{f}$. This leads to

$$\begin{aligned} A^N \widetilde{f}(\eta, \zeta) =& A^N_{neu} f(\eta) + \widehat{S} A^N_{sel} f(\eta) + \frac{1}{\hat{S}} A^N_{neu} f_1(\eta, \zeta) + A^N_{sel}(f_1(\eta, \xi)) - \widehat{S} f_1(\eta, \xi) \\ =& A^N_{neu} f(\eta) + A^N_{sel}(f_1(\eta, \xi)) + \mathcal{O}\left(\frac{1}{\widehat{S}} + \frac{\widehat{S}}{S} + \frac{\widehat{S}}{K} \right) \end{aligned}$$

where we have used (4.42), that f does not depend on ξ (which allows us to observe that $A_{env} f(\eta) = \mathbb{E}_\pi[f(\eta)] - f(\eta) = 0$) and (4.34), (which allows us to observe that as $\mathbb{E}_\pi[f_1(\eta, \zeta)] = 0$, $A_{env} f_1(\eta, \zeta) = -f_1(\eta, \zeta)$). Therefore, at least heuristically, the identification of $A^N_{sel}(f_1(\eta, \xi))$ should lead to the correct limit.

To make this argument rigorous, we use Theorem 3.16. We observe that

$$\begin{aligned} f(\eta_t) - \int_0^t A f(\eta_t, \zeta_t) \mathrm{d}s + (\widehat{f}(\eta_t, \zeta_t) - f(\eta_t)) + \int_0^t A f(\eta_s, \zeta_s) - A^N \widehat{f}(\eta_s, \zeta_s) \mathrm{d}s \\ = \widetilde{f}(\eta_t, \zeta_t) - \int_0^t A^N \widetilde{f}(\eta_s, \zeta_s) \mathrm{d}s, \end{aligned}$$

where A is given by (4.37). Since $\widetilde{f} - \int_0^t A^N \widetilde{f}(s)\mathrm{d}s$ is a martingale, we have written our problem in the form (3.20) with

$$\epsilon_N^f(t) = (\widetilde{f}(\eta_t, \zeta_t) - f(\eta_t)) + \int_0^t Af(\eta_s, \zeta_s) - A^N \widetilde{f}(\eta_s, \zeta_s)\mathrm{d}s \tag{4.43}$$

To check that the assumptions of Theorem 3.16 are satisfied, we work with both the lookdown representation and the projected model. The projected model allows us to check the compact containment condition (3.19) and prove the L^p estimate (3.21). Both are achieved via an intensity estimate given by the following Lemma.

Lemma 4.24. *Let $X^N = Kw^N$ denote the total intensity of individuals of the rare type. Assume that $\mathbb{E}[X^N(0)] < \infty$. Then for any $T > 0$*

$$\sup_{t \le T} \sup_N \mathbb{E}[X^N(t)] < \infty, \tag{4.44}$$

$$\lim_{H \to \infty} \sup_N \mathbb{P}\left[\sup_{t \le T} X^N(t) > H\right] = 0. \tag{4.45}$$

We discuss the proof in Subsection 4.4.3.1.

The part of the argument which allows us to identify the correct limit and shows that condition (3.22) is satisfied, that is

$$\lim_{n \to \infty} \mathbb{E}\left[\sup_{t \le T} |\epsilon_N^f(t)|\right] = 0, \tag{4.46}$$

is more involved, and requires the use of the lookdown representation. We will first look at the behaviour of

$$\int_0^t A^N f(\eta_s, \zeta_s) - Af(\eta_s, \zeta_s)\mathrm{d}s.$$

The terms involving A_{neu}^N and A_{sel}^N are tackled separately.

Let η_t^r be the process obtained from η_t by only considering the individuals of the rare type, $\eta_t^r := \{l : (l, \kappa_r) \in \eta_t\}$.

Proposition 4.25. *Under the conditions of Theorem 4.20,*

$$\mathbb{E}\Bigg[\sup_{t\leq T}\Bigg|\int_0^t A^N_{neu}f(\eta_s) - \left(f(\eta^r_s)\sum_{l_i(t)\in\eta^N_t} al_i^2\frac{g'(l_i(t))}{g(l_i(t))} + 2af(\eta^r_s)\sum_{l_i(t)\in\eta^r_s}\int_{l_i(t)}^{\infty}(1-g(\kappa_i,v))\mathrm{d}v\right)\mathrm{d}s\Bigg|\Bigg] \to 0.$$

This proposition will be proved via Taylor's formula through Proposition 4.31 and Proposition 4.32 in Subsection 4.4.3.2. An analogous proposition applies to the terms involving A^N_{sel}.

Proposition 4.26. *Under the conditions of Theorem 4.21 for any* $T \in \mathbb{R}$

$$\mathbb{E}\left[\sup_{t\leq T}\left|\int_0^t A^N_{sel}f(\eta^N_s) - f(\eta^N_s)\left(\sum_{l_i(t)\in\eta^N_t} -bl_i\frac{g'(l_i)}{g(l_i)}\right)\mathrm{d}s\right|\right] \to 0.$$

We discuss the proof of this proposition, along with calculations which allow us to fully justify the separation of timescales procedure in Subsection 4.4.3.3. These three propositions will allow us to conclude Theorem 4.20 and Theorem 4.21. We then proceed to use these results to prove Theorem 4.18 in Subsection 4.4.3.4. We notice that if the random perturbation Y appearing in the statement of Theorem 3.16 is trivial, the statement itself reduces to the usual condition for relative compactness of the sequence of stochastic processes, see, for example, Ethier and Kurtz (1986), Theorem 3.9.1 and Theorem 3.9.4. Therefore as a by-product of our construction we give a proof of Theorem 4.20 and Theorem 4.21. The proof of Theorem 4.20 is the most technical of this section.

Remark 4.27. *Our proof guarantees the relative compactness of the sequences of scaled lookdown representations and that limit points must satisfy a martingale problem. However, we do not have a proof of uniqueness of the martingale problem characterizing the limiting equation. We will use the Markov Mapping Theorem to deduce the relative compactness of the sequence of projected models and a martingale problem characterising limit points. Lemma A.13 from Kurtz and Rodrigues (2011) guarantees that every projection given by the Markov Map solves the projected martingale problem. Since this projected martingale problem has unique solutions, the relative compactness is enough*

to guarantee convergence of the sequence of projected models. This will not be the case in Section 4.5.

Remark 4.28. *We are interested in the behaviour of a rare subpopulation. As we have discussed earlier, we would like it to form $\mathcal{O}(1/K)$ of the population. For technical reasons, instead of the process $X^N(t)$, it is sometimes convenient to consider a stopped process $X^N(t \wedge \tau^N)$, where*

$$\tau^N := \inf\{t > 0 : X_t^N > Z^N\}.$$

We require the sequence of real numbers Z^N to be finite for each N and to tend to infinity as N tends to infinity. This requirement coupled with Lemma 4.24 guarantees that the convergence of the stopped processes translates directly into convergence of the unstopped processes. We shall see that we require that $Z^N \to \infty$ sufficiently slowly.

4.4.3.1 Intensity estimate

This subsection is devoted to the proof of our intensity estimate.

Proof[Proof of Lemma 4.24] The generator of the projected process can be written as

$$\begin{aligned}\mathcal{L}f(w,\zeta) = &\{wf((1-u)w+u,\zeta) + (1-p)f((1-u)w,\zeta) - f(w,\zeta)\}\\ &+ s\Big[\frac{\sigma(\kappa_r,\zeta)w}{\sigma(\kappa_r,\zeta)w + \sigma(\kappa_s,\zeta)(1-w)} pf((1-u)w+u,\zeta)\\ &+ \frac{\sigma(\kappa_c,\zeta)(1-w)}{\sigma(\kappa_r,\zeta)w + \sigma(\kappa_s,\zeta)(1-w)} wf((1-u)w+u,\zeta) - f(w,\zeta)\Big]\\ &+ \mathcal{L}^{env} f(p,\zeta).\end{aligned}$$

Recall that we are interested in an estimate for the intensity of the process describing the evolution of the rare individuals. We therefore substitute $X = Kw$. Under the assumptions of Theorem 4.18 this leads to the generator

$$\begin{aligned}\mathcal{L}f(X,\zeta) = N\Big\{\frac{X}{K} f\Big(\Big(1-\frac{u}{J}\Big)X + K\frac{u}{J},\zeta\Big) + \Big(1-\frac{X}{K}\Big) f\Big(\Big(1-\frac{u}{J}\Big)X,\zeta\Big) - f(X,\zeta)\Big\}\\ + N\widehat{S}\frac{s}{S}\Big[\frac{\sigma(\kappa_r,\zeta)\frac{X}{K}}{\sigma(\kappa_r,\zeta)\frac{X}{K} + \sigma(\kappa_s,\zeta)\Big(1-\frac{X}{K}\Big)} f\Big(\Big(1-\frac{u}{J}\Big)X + K\frac{u}{J},\zeta\Big)\end{aligned}$$

$$+\frac{\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_s,\zeta)\Big(1-\frac{X}{K}\Big)}f\bigg(\Big(1-\frac{u}{J}\Big)X,\zeta\bigg)-f(X,\zeta)\Bigg]$$
$$+\mathcal{L}^{env}f(X,\zeta).$$

This means that for any $f\in C^\infty(\mathbb{R})$,

$$f(X^N(T),\zeta_T)=f(X_0,\zeta_0)+\int_0^T\mathcal{L}(X^N(s),\zeta_s)\mathrm{d}s+M(T),$$

where M is a martingale. Substituting $f(x,\zeta)=x$ leads to

$$\begin{aligned}X(T)=&X(0)+N\widehat{S}\frac{s}{S}\frac{u}{J}\int_0^T\Bigg[\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}\frac{X}{K}(K-X)\\&+\frac{-\sigma(\kappa_c,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}\Big(1-\frac{X}{K}\Big)X\Bigg]\mathrm{d}s+M(T)\\=&X(0)+M(T)+N\widehat{S}\frac{s}{S}\frac{u}{J}\int_0^T\Bigg[\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}X\\&-\bigg(\frac{\sigma(\kappa_r,\zeta)\frac{X}{K}}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}+\frac{\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}\bigg)X\Bigg]\mathrm{d}s.\end{aligned}$$

Since

$$\begin{aligned}&\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}=\\&\qquad\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}+\frac{X}{K}\frac{\sigma(\kappa_c,\zeta)\sigma(\kappa_r,\zeta)-\sigma^2(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)\Big(\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)\Big)}\end{aligned}\tag{4.47}$$

this expression can be written as

$$\begin{aligned}X(T)=X(0)+N\widehat{S}\frac{s}{S}\frac{u}{J}\Bigg\{&\int_0^T\bigg[\Big(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}-1\Big)X\bigg]\mathrm{d}s\\&+\int_0^T\Bigg[\frac{X^2}{K}\frac{\sigma(\kappa_c,\zeta)\sigma(\kappa_r,\zeta)-\sigma^2(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)\Big(\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)\Big)}\Bigg]\mathrm{d}s\Bigg\}+M(T).\end{aligned}\tag{4.48}$$

Consider the stopping time $\tau=\inf\{t\geq 0: X^N>H\}$ and the stopped process $\widehat{X}^N(t)=X^N(t\wedge\tau)$. Since (4.48) holds at a bounded stopping time, taking the

expectation and using the symmetry condition (4.34) leads to

$$\mathbb{E}[\widehat{X}^N(T)] = \mathbb{E}[\widehat{X}(0)] + N\widehat{S}\frac{s}{S}\frac{u}{J}\mathbb{E}\Bigg[\int_0^T\Big\{\Big(\frac{\sigma(\kappa_c,\zeta)}{\sigma(\kappa_r,\zeta)}-1\Big)\widehat{X}^N(s)\Big\}\mathrm{d}s$$
$$+\int_0^T\Bigg\{\frac{(\widehat{X}^N(s))^2}{K}\frac{\sigma^2(\kappa_r,\zeta)+\sigma(\kappa_c,\zeta)\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)\Big(\sigma(\kappa_r,\zeta)\frac{\widehat{X}^N(s)}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{\widehat{X}^N(s)}{K}\Big)\Big)}\Bigg\}\mathrm{d}s\Bigg]$$
$$\leq \mathbb{E}[\widehat{X}(0)] + N\frac{H}{K}\frac{s\widehat{S}}{S}\frac{u}{J}C_\sigma\int_0^T\mathbb{E}[\widehat{X}^N(s)]\mathrm{d}s,$$

where C_σ is a constant depending on σ. By Grönwall's inequality

$$\mathbb{E}[\widehat{X}^N(T)] \leq \mathbb{E}[\widehat{X}^N(0)]\exp\Big(N\frac{H}{K}\frac{s\widehat{S}}{S}\frac{u}{J}C_\sigma T\Big).$$

If one takes $H = Z^N$ this, combined with the conditions from Theorem 4.18, concludes the proof of (4.44). To see that (4.45) holds, it is enough to observe that by Markov's inequality

$$\mathbb{P}\Big[\sup_{0\leq t\leq T}X_t \geq H\Big] = \mathbb{P}[\widehat{X}_T \geq H] \leq \frac{\mathbb{E}[\widehat{X}_T]}{H}, \tag{4.49}$$

and for any T, $\varepsilon > 0$ we can choose H_1, N_1 such that for $H \geq H_1$, $N \geq N_1$ and $t \leq T$ the right hand side of (4.49) is less than ε. □

We also take note of a simple corollary which we will use later in our main proof.

Corollary 4.29. *The lookdown process satisfies the compact containment condition,* (3.19) *in Theorem 3.16.*

Proof We note that the characterisation of convergence given in Theorem 2.92 ensures that a sequence of lookdown processes satisfies the compact containment condition if and only if the projected processes also satisfy compact containment condition. □

Remark 4.30. *We do not give details but a similar calculation also proves Lemma 4.24 under the assumptions of Theorem 4.20 and Theorem 4.21.*

4.4.3.2 Neutral model - proof of Theorem 4.20

Even though this subsection contains the proof of the result for the least complicated model, the proof itself is the most involved one. The relative simplicity of the proof for the more complicated model demonstrates the power of this lookdown method. Proofs of Theorem 4.21 and Theorem 4.18 heavily rely on technical observations from this subsection. We recall that the generator of the neutral part of the process takes the form

$$A^N_{neu}f(\eta) = N\Bigg(\int_0^\infty \Bigg[\frac{uK}{J}e^{-\frac{uK}{J}v^*}g(\kappa^*,v^*)e^{-\frac{uK}{J}\int_{v^*}^\infty(1-g(\kappa^*,v))\mathrm{d}v} \times \prod_{(\kappa,l)\in\eta,l\neq l^*} g(\kappa,\mathcal{J}_{neu}(l,l^*,v^*))\Bigg]\mathrm{d}v^* - f(\eta)\Bigg), \quad (4.50)$$

where (κ^*,v^*) denotes the parent and $\mathcal{J}_{neu}$ is defined as in (4.31), that is

$$\mathcal{J}_{neu}(l,l^*_{neu},v^*) = \begin{cases} \frac{1}{1-u}(l-(l^*_{neu}-v^*)) & \text{if } l > l^*_{neu}, \\ \frac{1}{1-u}l & \text{if } l < l^*_{neu}, \\ v^* & \text{if } l = l^*_{neu}. \end{cases}$$

Before stating the propositions and lemmas which prove Theorem 4.20, let us rewrite (4.50) in a more convenient form.

We start by observing that Taylor's expansion of g gives

$$\begin{aligned} &g(\kappa^*,v^*)\prod_{(\kappa,l)\in\eta,l\neq l^*} g(\kappa,\mathcal{J}(l,l^*,v^*)) - f(\eta) \\ &= f(\eta)\sum_l \frac{g'(\kappa,l)}{g(\kappa,l)}(\mathcal{J}(l,l^*,v^*)-l) + \sum_{l,\hat{l}}\frac{g'(\hat{\kappa},\hat{l})}{g(\hat{\kappa},\hat{l})}\frac{g'(\kappa,l)}{g(\kappa,l)}\mathcal{O}((\mathcal{J}(l,l^*,v^*)-l)(\mathcal{J}(\hat{l},l^*,v^*)-\hat{l})) \\ &\quad + \sum_l \frac{g''(\kappa,l)}{g^2(\kappa,l)}\mathcal{O}((\mathcal{J}^2(l,l^*,v^*)-l)), \quad (4.51) \end{aligned}$$

and a Taylor expansion of the exponential function about 0 leads to

$$e^{-\frac{uK}{J}\int_{v^*}^\infty(1-g(\kappa^*,v))\mathrm{d}v} =$$

$$1-\frac{uK}{J}\int_{v^*}^{\infty}(1-g(\kappa^*,v))\mathrm{d}v+\mathcal{O}\Big(\frac{uK}{J}\int_{v^*}^{\infty}(1-g(\kappa^*,v))\mathrm{d}v\Big)^2. \quad (4.52)$$

Applying (4.52) we may rewrite (4.50) as

$$A^N_{neu}=A^N_{neu,1}+A^N_{neu,2}+\mathcal{O}\Big(\frac{K}{J}A^N_{neu,2}\Big),$$

where

$$A^N_{neu,1}=N\bigg(\int_0^{\infty}\frac{uK}{J}e^{-\frac{uK}{J}v^*}\Big[g(\kappa^*,v^*)\prod_{(\kappa,l)\in\eta,l\neq l^*}g(\kappa,\mathcal{J}(l,l^*,v^*))-f(\eta)\Big]\mathrm{d}v^*\bigg), \quad (4.53)$$

$$A^N_{neu,2}=N\bigg(-\int_0^{\infty}\frac{u^2K^2}{J^2}e^{-\frac{uK}{J}v^*}\int_{v^*}^{\infty}(1-g(\kappa^*,v))\mathrm{d}v\; g(\kappa^*,v^*)\times\prod_{(\kappa,l)\in\eta,l\neq l^*}g(\kappa,\mathcal{J}(l,l^*,v^*))\mathrm{d}v^*\bigg). \quad (4.54)$$

We begin with statements of the propositions that identify limits of (4.53) and (4.54) separately. Their proofs are later in this section.

Proposition 4.31. *Under the conditions of Theorem 4.20,*

$$\mathbb{E}\bigg[\sup_{t\leq T}\bigg|\int_0^t A^N_{neu,1}f(\eta_s)-f(\eta^r_s)\bigg(\sum_{l_i(t)\in\eta^r_t}al_i^2\frac{g'(l_i(t))}{g(l_i(t))}\bigg)\mathrm{d}s\bigg|\bigg]\to 0.$$

Proposition 4.32. *Under the conditions of Theorem 4.20,*

$$\mathbb{E}\bigg[\sup_{t\leq T}\bigg|\int_0^t\bigg\{A^N_{neu,2}f(\eta_s)-2af(\eta^r_s)\sum_{l_i(t)\in\eta^r_s}\int_{l_i(t)}^{\infty}(1-g(\kappa_i,v))\mathrm{d}v\bigg\}\mathrm{d}s\bigg|\bigg]\to 0.$$

The proof of Proposition 4.31 will use the following elementary lemma:

Lemma 4.33. *Let Γ be a stochastic process. Assume that the second moment of Γ is bounded uniformly for all times up to time T by ϵ, that is $\mathbb{E}[(\Gamma(s))^2]\leq\epsilon$, for $0\leq s\leq T$. Then*

$$\mathbb{E}\bigg[\sup_{t<T}\bigg|\int_0^t\Gamma(s)\mathrm{d}s\bigg|\bigg]<\sqrt{T\epsilon}.$$

Proof By Jensen's inequality

$$\mathbb{E}\left[\sup_{t\leq T}\left(\int_0^t \Gamma(s)\mathrm{d}s\right)^2\right] \leq \mathbb{E}\left[\sup_{t\leq T} t\int_0^t \Gamma^2(s)\mathrm{d}s\right] = T\mathbb{E}\left[\int_0^T \Gamma^2(s)\mathrm{d}s\right].$$

The inequality follows by assumption and a final application of Jensen's inequality. □

Proof[Proof of Proposition 4.31] By Lemma 4.33, it suffices to prove that, conditioned on the postion of rare levels,

$$\mathbb{E}\Bigg[\Bigg\{N\Bigg(\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}\Bigg\{g(\kappa^*,v^*) \times \prod_{(\kappa(t),l(t))\in\eta, l(t)\neq l^*(t)} g(\kappa,\mathcal{J}(l(t),l^*(t),v^*)) - f(\eta_s)\Bigg\}\mathrm{d}v^*\Bigg) - f(\eta_s^r)\left(\sum_i a l_i^2\frac{g'(l_i(t))}{g(l_i(t))}\right)\Bigg\}^2\Bigg] \to 0,$$

uniformly for $0 \leq s \leq T$.

For ease of notation we let $l_0 = 0$. We will use the ordering l_i and (4.51) to observe that $A^N_{neu,1}$ can be approximated by

$$\begin{aligned}
&A^N_{neu,1}f(\eta)\\
&= Nf(\eta)\left(\sum_i\int_{l_{i-1}}^{l_i}\frac{uK}{J}e^{-\frac{uK}{J}v^*}\left[\frac{g'(\kappa_i,l_i)}{g(\kappa_i,l_i)}(v^*-l_i)+\sum_{l\neq l_i}\frac{g'(\kappa,l)}{g(\kappa,l)}\left(\frac{l\frac{u}{J}}{1-\frac{u}{J}}-\mathbf{1}_{l>l_i}\frac{l_i-v^*}{1-\frac{u}{J}}\right)\right]\mathrm{d}v^*\right)\\
&= Nf(\eta)\sum_i\Bigg[\frac{g'(\kappa_i,l_i)}{g(\kappa_i,l_i)}\left(-(l_i-l_{i-1})e^{-\frac{uK}{J}l_{i-1}}+\frac{J}{uK}\left(e^{-\frac{uK}{J}l_{i-1}}-e^{-\frac{uK}{J}l_i}\right)\right)\\
&\qquad+\sum_{j\neq i}\left(\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\frac{l_j\frac{u}{J}}{1-\frac{u}{J}}\left(e^{-\frac{uK}{J}l_{i-1}}-e^{-\frac{uK}{J}l_i}\right)\right)\\
&+\sum_{j\neq i}\sum_{j<i}\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\frac{1}{1-\frac{u}{J}}\left(-(l_i-l_{i-1})e^{-\frac{uK}{J}l_{i-1}}+\frac{J}{uK}\left(e^{-\frac{uK}{J}l_{i-1}}-e^{-\frac{uK}{J}l_i}\right)\right)\Bigg], \quad (4.55)
\end{aligned}$$

where the second line follows from integration. At this stage we note that we can drop the factor $1-u/J$ at the cost of an error of order NK/J^3, which tends to zero as N tends to infinity. We also swap the order of summation to rewrite (4.55) as

$$A^N_{neu,1}f(\eta) = Nf(\eta)\sum_j\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}$$

$$\times\left[\sum_{i\leq j}\left(-(l_i-l_{i-1})e^{-\frac{uK}{J}l_{i-1}}+\frac{J}{uK}\Big(e^{-\frac{uK}{J}l_{i-1}}-e^{-\frac{uK}{J}l_i}\Big)\right)+l_j\frac{u}{J}\sum_{i\neq j}\Big(e^{-\frac{uK}{J}l_{i-1}}-e^{-\frac{uK}{J}l_i}\Big)\right].$$

We observe that since the second and the third terms in the inner sum are telescoping sums and $l_0=0$ we simplify our expression to

$$\begin{aligned}A^N_{neu,1}f(\eta)=Nf(\eta)\sum_j\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\Bigg[&\frac{J}{uK}\Big(1-e^{-\frac{uK}{J}l_j}\Big)+l_j\frac{u}{J}\Big(1-e^{-\frac{uK}{J}l_{j-1}}+e^{-\frac{uK}{J}l_j}\Big)\\&+\sum_{i\leq j}\Big(-(l_i-l_{i-1})e^{-\frac{uK}{J}l_{i-1}}\Big)\Bigg]\\&=:f(\eta)\sum_j\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}(P_1+P_2+P_3).\quad(4.56)\end{aligned}$$

Now we treat each term in the new sum separately. Using the Taylor expansion of the exponential function about 0 we observe that

$$\begin{aligned}P_1&=\frac{NJ}{uK}\Big(1-e^{-\frac{uK}{J}l_j}\Big)=N\sum_{1\leq k\leq n+1}\frac{(-1)^{k-1}}{k!}l_j^k\Big(\frac{uK}{J}\Big)^{k-1}+\mathcal{O}\Big(\frac{N(uK)^{n+1}}{J^{n+1}}\Big),\\P_2&=l_j\frac{uN}{J}\Big(1-e^{-\frac{uK}{J}l_{j-1}}+e^{-\frac{uK}{J}l_j}\Big)=\frac{uNl_j}{J}+\mathcal{O}\Big(\frac{NK}{J^2}(l_j-l_{j-1})\Big),\\P_3&=N\sum_{i\leq j}\Big(-(l_i-l_{i-1})e^{-\frac{uK}{J}l_{i-1}}\Big)=N\sum_{1\leq k\leq n+1}\frac{(-1)^k}{(k-1)!}\Big(\frac{uK}{J}\Big)^{k-1}\sum_{i\leq j}(l_i-l_{i-1})l_{i-1}^{k-1}\\&\qquad+\mathcal{O}\Big(\frac{N(uK)^{n+1}}{J^{n+1}}\Big).\end{aligned}$$

We focus our attention on P_3. We investigate the terms corresponding to different values of k separately. We observe that the first three terms involve

$$\sum_{i\leq j}(l_i-l_{i-1})=l_j,\quad(4.57)$$

$$\begin{aligned}\sum_{i\leq j}(l_i-l_{i-1})l_{i-1}&=\frac{1}{2}\left(\sum_{i\leq j}(l_i^2-l_{i-1}^2)-\sum_{i\leq j}(l_i-l_{i-1})^2\right)\\&=\frac{1}{2}\left(l_j^2-\sum_{i\leq j}(l_i-l_{i-1})^2\right)\quad(4.58)\end{aligned}$$

$$\begin{aligned}\sum_{i\leq j}\Big((l_i-l_{i-1})l_{i-1}^2\Big)&=\frac{1}{3}\sum_{i\leq j}\Big(l_i^3-l_{i-1}^3\Big)-\frac{1}{3}\sum_{i\leq j}(l_i-l_{i-1})^3-\sum_{i\leq j}(l_i-l_{i-1})^2l_{i-1}\\&=\frac{1}{3}l_j^3-\frac{1}{3}\sum_{i\leq j}(l_i-l_{i-1})^3-\sum_{i\leq j}(l_i-l_{i-1})^2l_{i-1},\quad(4.59)\end{aligned}$$

respectively. We are therefore interested in $\sum_i (l_i - l_{i-1})^2$ conditioned on the locations of the rare levels, that is, conditioned on η^r.

Recall that conditioned on the locations of the levels of the rare type, the levels of the common type will be Poisson distributed with intensity $K - Z^N$. Conditioned on the number of levels of the rare type between 0 and l_j these levels will be uniformly distributed on $[0, l_j]$. We denote n independent uniformly distributed random variables on $[0,1]$ by $u_1, \ldots, u_n$ and define the order statistics by letting $u_{(i)} = u_k$ if and only if $\#\{\hat{k} : u_{\hat{k}} \leq u_k\} = i$ for $i \in \{1, \ldots n\}$. We note that n points uniformly distributed on $[0,1]$ can be identified with $n+1$ points uniformly distributed on the unit circle with one of these points chosen at random to be a reference point corresponding to both 0 and 1. This then leads us to see that $u_{(i)} - u_{(i-1)}$ is equal in distribution to $u_{(1)}$ for $i \in \{1, \ldots, n+1\}$, where $u_{(n+1)} := 1$ and $u_{(0)} := 0$. From this one can see that

$$\mathbb{E}\left[\sum_{i=1}^{j}(l_i - l_{i-1})^2 | l_0 = 0, j = n+1, l_j\right] = l_j^2 \mathbb{E}\left[\sum_{i=1}^{n+1}(u_{(i)} - u_{(i-1)})^2\right] = l_j^2 (n+1)\frac{2}{(n+1)(n+2)}.$$

We then use that the number of levels of the rare type within $[0, l_j]$ will be Poisson distributed to see

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{j}(l_i - l_{i-1})^2 | l_0 = 0, l_j\right] =& \sum_{i=0}^{\infty}\frac{2l_j^2}{n+2}\frac{\Big(l_j(K-Z^N)\Big)^n \exp(-\Big(l_j(K-Z^N)\Big))}{n!} \\ =& \frac{2l_j}{(K-Z^N)} + \mathcal{O}\Big(\exp\big(-l_j(K-Z^N)\big)\Big).\end{aligned}$$

Identical calculations show

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^{j}(l_i - l_{i-1})^2\right)^2 | l_0 = 0, l_j\right] =& \frac{4l_j}{(K-Z^N)^3}(3 + x(K-Z^N)) \\ &+ \mathcal{O}\Big((K-Z^N)\exp\big(-l_j(K-Z^N)\big)\Big), \qquad (4.60) \\ \mathrm{Var}\left(\sum_{i=1}^{j}(l_i - l_{i-1})^2 | l_0 = 0, l_j\right) =& \mathcal{O}\Big(\frac{l_j}{(K-Z^N)^3}\Big).\end{aligned}$$

We note that we are considering j to be a random variable throughout this corresponding to the level of a given rare individual.

From this calculation since we multiply (4.58) by NK/J in we see that we require $\frac{N^2}{KJ^2} \to 0$. Therefore, we may approximate $\left(\frac{uK}{J}\right)\sum_{i\leq j}(l_i - l_{i-1})^2$ by $\frac{ul_j}{J}$. For $k = 3$ we again condition on the number of levels of the rare type beneath l_j and see that

$$\mathbb{E}\left[\sum_{i=1}^{j}(l_i - l_{i-1})^2 l_{i-1} | l_0 = 0, j = n+1, l_j\right] = (n+1)\mathbb{E}\Big[(l_I - l_{I-1})^2 l_{I-1} | l_0 = 0, j = n+1, l_j\Big],$$

where I is chosen uniformly at random from $(1, \cdots, n+1)$. We then again consider the levels as $n+1$ points chosen uniformly at random from a circle to see that this will be $\frac{l_j^3}{n+2}$. This then gives us

$$\mathbb{E}\left[\sum_{i\leq j}(l_i - l_{i-1})^2 l_{i-1} | l_0 = 0, l_j\right] = \frac{l_j^2}{K - Z^N} + \mathcal{O}\Big(\exp(-l_j(K - Z^N))\Big). \tag{4.61}$$

We see that for $k \geq 3$

$$\mathbb{E}\left[\sum_{i\leq j}\Big((l_i - l_{i-1})l_{i-1}^k\Big) - \frac{1}{k+1}l_j^{k+1}\right] = \mathcal{O}\left(\frac{1}{K}\right), \tag{4.62}$$

$$\mathrm{Var}\left(\sum_{i\leq j}\Big((l_i - l_{i-1})l_{i-1}^k\Big) - \frac{1}{k+1}l_j^{k+1}\right) = \mathcal{O}\left(\frac{1}{K^3}\right), \tag{4.63}$$

which will suffice as each of these terms will be multiplied by $\frac{NK^{k-1}}{J^{k-1}}$ in (4.56). We then note that $\frac{NK}{J^2}$ is bounded and $\frac{K}{J} \to 0$. We combine (4.57), (4.58), (4.59), (4.62) to see that, conditioned on l_j,

$$\mathbb{E}\left[\sum_{i\leq j}\Big(-(l_i - l_{i-1})e^{-\frac{uK}{J}l_{i-1}}\Big)\right]$$
$$= \sum_{0\leq k\leq n}\frac{(-1)^{k+1}}{(k+1)!}\left(\frac{uK}{J}\right)^k l_j^{k+1} - \frac{u}{J}l_j + \frac{1}{2}\frac{u^2K}{J^2}l_j^2 + \mathcal{O}\left(\frac{K^2}{J^3}\right) + \mathcal{O}\left(\frac{K}{J}\right)^{n+1}.$$

Finally, we observe that all these approximations and cancellations allow us to approximate (4.56) by

$$f(\eta)\sum_j \frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\left(\frac{1}{2}\frac{u^2NK}{J^2}l_j^2+\mathcal{O}\left(\frac{NK^2}{J^3}\right)+\mathcal{O}\left(\frac{NK^{n+1}}{J^{n+1}}\right)+\mathcal{O}\Big(\frac{NK}{J^2}(l_j-l_{j-1})\Big)\right),$$

plus a random, mean zero, correction term with variance $\mathcal{O}(N^2K^2/(J^2(K-Z)^3))$. □

The proof of Proposition 4.32 is more involved than that of Proposition 4.31. Once again we begin with an elementary lemma.

Lemma 4.34. *Let $\Gamma(s)$ be a stochastic process. For any partition $0=t_0<t_1<\cdots<t_m=T$,*

$$\mathbb{E}\left[\sup_{t\le T}\left|\int_0^t\Gamma(s)\mathrm{d}s\right|\right]\le\sum_{j=1}^m\sqrt{\mathbb{E}\left[\left(\int_{t_{j-1}}^{t_j}\Gamma(s)\mathrm{d}s\right)^2\right]}+\mathbb{E}\left[\sup_j\int_{t_j-1}^{t_j}|\Gamma(s)|\mathrm{d}s\right].$$

Proof We observe that

$$\begin{aligned}\mathbb{E}\left[\sup_{t\le T}\left|\int_0^t\Gamma(s)\mathrm{d}s\right|\right]&\le\mathbb{E}\left[\sum_{j=1}^m\left|\int_{t_{j-1}}^{t_j}\Gamma(s)\mathrm{d}s\right|+\sup_j\int_{t_j-1}^{t_j}|\Gamma(s)|\mathrm{d}s\right]\\&\le\sum_{j=1}^m\mathbb{E}\left[\left|\int_{t_{j-1}}^{t_j}\Gamma(s)\mathrm{d}s\right|\right]+\mathbb{E}\left[\sup_j\int_{t_j-1}^{t_j}|\Gamma(s)|\mathrm{d}s\right]\\&\le\sum_{j=1}^m\sqrt{\mathbb{E}\left[\left(\int_{t_{j-1}}^{t_j}\Gamma(s)\mathrm{d}s\right)^2\right]}+\mathbb{E}\left[\sup_j\int_{t_j-1}^{t_j}|\Gamma(s)|\mathrm{d}s\right].\end{aligned}$$

which concludes the proof. □

We require a few more computations to transform $A^N_{neu,2}$ into a more convenient form. First of all, we observe that as $(1-g)$ is bounded, the following approximation is valid:

$$\begin{aligned}\int_{l_{i-1}}^{l_i}\int_{v^*}^{\infty}(1-g(\kappa_i,v))\mathrm{d}v\,\mathrm{d}v^*&=\int_{l_{i-1}}^{\infty}\int_{l+i-1}^{l_i\wedge v}(1-g(\kappa_i,v))\mathrm{d}v^*\,\mathrm{d}v\\&=\int_{l_i}^{\infty}\int_{l+i-1}^{l_i}(1-g(\kappa_i,v))\mathrm{d}v^*\,\mathrm{d}v+\int_{l_{i-1}}^{l_i}\int_{l+i-1}^{v}(1-g(\kappa_i,v))\mathrm{d}v^*\,\mathrm{d}v\\&=(l_i-l_{i-1})\int_{l_i}^{\infty}(1-g(\kappa_i,v))\mathrm{d}v+\mathcal{O}\Big((l_i-l_{i-1})^2\Big).\end{aligned}\quad(4.64)$$

For convenience, we order the individuals present in the system according to their level, that is we consider $\eta = \{(\kappa_i, l_i)\}_{i\geq 1}$ where $l_i < l_{i+1}$. We observe that the ordering leads to the following simplification:

$$A^N_{neu,2} f(\eta) = -\frac{u^2 NK}{J} \sum_i \left(\int_{l_{i-1}}^{l_i} \frac{uK}{J} e^{-\frac{uK}{J} v^*} f(\eta) \int_{v^*}^{\infty} (1 - g(\kappa_i, v)) \mathrm{d}v \, \mathrm{d}v^* \right).$$

Since $\frac{K}{J} \to 0$, we may use (4.64) to further simplify $A^N_{neu,2}$ to

$$\begin{aligned} &A^N_{neu,2} f(\eta) \\ &\quad = -\frac{u^2 NK^2}{J^2} f(\eta) \sum_i (l_i - l_{i-1}) \left(\int_{l_i}^{\infty} (1 - g(\kappa_i, v)) \mathrm{d}v \, \mathrm{d}v^* \right) \left(1 + \mathcal{O}\left(\frac{K}{J} + \frac{1}{K^2} \right) \right). \end{aligned}$$

By this calculation and Lemma 4.34, it is clear that in order to prove Proposition 4.32, it is enough to show that for any partition $0 = t_0 < t_1 < \cdots < t_m = T$,

$$\sum_{j=1}^{m} \sqrt{\mathbb{E}\left[\left(\int_{t_{j-1}}^{t_j} \Gamma^N(s) \mathrm{d}s \right)^2 \right]} + \mathbb{E}\left[\sup_j \int_{t_j - 1}^{t_j} |\Gamma^N(s)| \mathrm{d}s \right] \to 0, \tag{4.65}$$

where

$$\Gamma^N(s) = \sum_i \left(\frac{u^2 NK^2}{J^2} (l_i(t) - l_{i-1}(t)) - 2a \right) f(\eta_t) \int_{l_i(t)}^{\infty} (1 - g(\kappa_i, v)) \mathrm{d}v \, \mathrm{d}v^* \mathrm{d}t. \tag{4.66}$$

We first turn our attention to the parts without a supremum which we can calculate directly.

Lemma 4.35. *Conditioned on the locations of the rare levels,*

$$\begin{aligned} \mathbb{E}\left[\left(\int_0^T \sum_i \left(\frac{u^2 NK^2}{J^2} (l_i(t) - l_{i-1}(t)) - 2a \right) f(\eta_t) \int_{l_i(t)}^{\infty} (1 - g(\kappa_i, v)) \mathrm{d}v \, \mathrm{d}v^* \mathrm{d}t \right)^2 \right] \\ = \mathcal{O}\left(\frac{LT}{N} \right) + o(T^2), \end{aligned} \tag{4.67}$$

for any $L(N)$ such that $\frac{L}{J} \to \infty$.

Proof We recall that there exists λ_g such that for all $v \geq \lambda_g$, $g(v) = 1$. Therefore

$$f(\eta_t) \int_{l_i(t)}^{\infty} (1 - g(\kappa_i, v)) \mathrm{d}v$$

is bounded for each test function. As η_t is constant between events we condition on exactly n reproductive events occurring within $[0, T]$. We denote the time of events by t_i for $i \in \{1, \ldots, n\}$. Observe that t_i are independent random variables uniformly distributed on $[0, T]$. For ease of notation we let $t_0 = 0$ and $t_{n+1} = T$. We will use $\hat{l}_k$ to denote a rare type individual's level and $\tilde{l}_k$ to be the highest level across the population (that is, from individuals of both rare and common type) below $\hat{l}_k$ and consider

$$\mathbb{E}\Bigg[\sum_{i=0}^{n}\sum_{j=0}^{n}(t_{i+1} - t_i)(t_{j+1} - t_j) \left(\frac{u^2NK^2}{J^2}(\hat{l}_k(t_i) - \tilde{l}_k(t_i)) - 2a\right)\left(\frac{uNK^2}{J^2}(\hat{l}_k(t_j) - \tilde{l}_k(t_j)) - 2a\right)\Bigg]. \quad (4.68)$$

We note that the times of events are independent of the level process and so we use

$$\mathbb{E}[(t_{i+1} - t_i)(t_{j+1} - t_j)] = \begin{cases} \frac{T^2}{(n+1)(n+2)} & \text{if } i \neq j, \\ \frac{2T^2}{(n+1)(n+2)} & \text{if } i = j. \end{cases}$$

We also use the Cauchy–Schwarz inequality to see

$$\mathbb{E}[(\hat{l}_k(t_i) - \tilde{l}_k(t_i))(\hat{l}_k(t_j) - \tilde{l}_k(t_j))] \leq \frac{2}{(K - Z^N)^2},$$

where we note that we are still looking at the stopped process .

We denote the intensity, after n events, of the levels of all individuals which were offspring in one of those n events by I_n. It is then easy to see $I_{n+1} = I_n(1 - \frac{u}{J}) + \frac{uK}{J}$ and $I_0 = 0$. This recurrence equation has a solution given by

$$I_n = K\left(1 - \left(1 - \frac{u}{J}\right)^n\right).$$

For $i < j$ we define $\tilde{l}_k^{new}(t_j)$ as the highest level below $\hat{l}_k$ of an individual (of either rare or common type) which has been born since t_i. Then we can see that

$$(\hat{l}_k(t_j) - \tilde{l}_k(t_j)) \leq (\hat{l}_k(t_j) - \tilde{l}_k^{new}(t_j)),$$

and so

$$\mathbb{E}[(\hat{l}_k(t_i) - \tilde{l}_k(t_i))(\hat{l}_k(t_j) - \tilde{l}_k(t_j))] \leq \frac{1}{K - Z^N}\frac{1}{I_{j-i}}.$$

We then also see that

$$\begin{aligned}
\mathbb{E}[(\hat{l}_k(t_i)-\tilde{l}_k(t_i))(\hat{l}_k(t_j)-\tilde{l}_k(t_j))] \geq& \mathbb{E}[(\hat{l}_k(t_i)-\tilde{l}_k(t_i))(\hat{l}_k(t_j)-\tilde{l}_k(t_j))\mathbf{1}_{\tilde{l}_k(t_j)=\tilde{l}_k^{new}(t_j)}] \\
=& \mathbb{E}[(\hat{l}_k(t_i) - \tilde{l}_k(t_i))(\hat{l}_k(t_j)-\tilde{l}_k^{new}(t_j))] \\
&-\mathbb{E}[(\hat{l}_k(t_i)-\tilde{l}_k(t_i))(\hat{l}_k(t_j)-\tilde{l}_k^{new}(t_j))\mathbf{1}_{\tilde{l}_k(t_j)\neq\tilde{l}_k^{new}(t_j)}] \\
\geq& \frac{1}{I_{j-i}}\frac{1}{K - Z^N} - \sqrt{\frac{2}{(K - Z^N)^2}\frac{2}{I_{j-i}^2}\mathbb{P}[\tilde{l}_k(t_j) \neq \tilde{l}_k^{new}(t_j)]},
\end{aligned}$$

where we have used the Cauchy-Schwartz inequality for the final line.

We note that

$$\mathbb{P}[\tilde{l}_k(t_j) \neq \tilde{l}_k^{new}(t_j)] \leq \frac{K(1 - \frac{u}{J})^{j-i}}{K(1 - \frac{u}{J})^{j-i} + I_{j-i}}.$$

We now consider a function $L(N)$. By splitting (4.68) into parts with $|i - j| \leq L$ and $|i - j| > L$, we bound it above by

$$\begin{aligned}
4(n+1)L\frac{T^2}{(n+1)(n+2)}&\left(\frac{N^2K^4}{J^4}\frac{2}{(K - Z^N)^2} + 4a^2\right) \\
&+ C(n+1)(n+1-L)\frac{T^2}{(n+1)(n+2)} \\
\times\Bigg\{&\frac{N^2K^4}{J^4}\frac{1}{K - Z^N}\frac{1}{I_L} - \left(\frac{NK^2}{J^2}\frac{1}{K - Z^N} + \frac{NK^2}{J^2}\frac{1}{I_L}\right)2a + 4a^2 \\
&+ \mathcal{O}\left(\frac{N^2K^4}{J^4}\frac{1}{K - Z^N}\frac{1}{I_L}\sqrt{\frac{K(1 - \frac{u}{J})^L}{K(1 - \frac{u}{J})^L + I_L}}\right)\Bigg\}.
\end{aligned}$$

We observe that

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{n+1}{(n+1)(n+2)}\frac{(NT)^n e^{-NT}}{n!} =& \frac{1}{NT}-\frac{1}{(NT)^2}+\frac{e^{-NT}}{(NT)^2}-\frac{e^{-NT}}{2},\\ \sum_{n=1}^{\infty}\frac{(n+1)^2}{(n+1)(n+2)}\frac{(NT)^n e^{-NT}}{n!} \leq& 1.\end{aligned}$$

We note that

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T\sum_i\left(\frac{u^2NK^2}{J^2}(l_i(t)-l_{i-1}(t))-2a\right)f(\eta_t)\int_{l_i(t)}^{\infty}(1-g(\kappa_i,v))\mathrm{d}v\,\mathrm{d}v^*\mathrm{d}t\right)^2\right]&\\ =\mathcal{O}\left(\frac{LT}{N}\left(\frac{N^2K^4}{J^4}\frac{2}{(K-Z)^2}+4a^2\right)\right.&\\ +T^2\left\{\frac{N^2K^4}{J^4}\frac{1}{K-Z}\frac{1}{I_L}-\left(\frac{NK^2}{J^2}\frac{1}{K-Z}+\frac{NK^2}{J^2}\frac{1}{I_L}\right)2a\right.&\\ \left.\left.+4a^2+\mathcal{O}\left(\frac{N^2K^4}{J^4}\frac{1}{K-Z}\frac{1}{I_L}\sqrt{\frac{K(1-\frac{u}{J})^L}{K(1-\frac{u}{J})^L+I_L}}\right)\right\}\right).&\end{aligned}$$

Therefore, in order to conclude that (4.67) converges to 0, it is enough to find L which satisfies

$$\frac{L}{N}\to 0;\quad \frac{NK^2}{J^2}\frac{1}{I_L}\to 2a;\quad \frac{K(1-\frac{u}{J})^L}{K(1-\frac{u}{J})^L+I_L}\to 0.$$

Since $\frac{N}{J}\to\infty$, this is achieved by any $L(N)$ such that $\frac{L}{J}\to\infty$ and $\frac{L}{N}\to 0$. for example, we can choose $L(N)=\sqrt{NJ}$.

□

We now turn our attention back to (4.65). We consider $t_j=j\delta(N)$ and see that to ensure the first part of (4.65) converges to zero, we need choices of L and δ such that

$$\frac{L}{N}\to 0;\quad \frac{L}{J}\to\infty;\quad \frac{L}{N\delta}\to 0. \tag{4.69}$$

To show that the second term of (4.65) converges to zero, it is enough to show that $\mathbb{E}[\sup_{t\leq T}\Gamma^N(t)]$ is bounded, provided that there exists a δ satisfying (4.69) such that $\delta\to 0$. The latter is satisfied by e.g. $\delta=\sqrt{L/N}$.

To show $\mathbb{E}[\sup_{t\leq T}\Gamma^N(t)]$ is bounded we consider an auxiliary process β^N, defined by

$$\beta_t^N = \sum_{l(t)\in\eta_t^r}\Big(l(t)-\hat{l}(t)\Big)h_\lambda(l(t)), \tag{4.70}$$

where h is a decreasing, positive cut-off function (that is, we assume that there exists a λ such that $h(v)=0$ for $v>\lambda$) and $\hat{l}(t)\in\eta_t$ is the first level below $l(t)$. Observe that to show the required bound on the expectation of the supremum of the integral of X, (and therefore conclude the proof), it is enough to show that β^N is dominated by a bounded submartingale.

Lemma 4.36. *Define β^N as in* (4.70)*. Then β^N is dominated by a bounded submartingale.*

Proof We recall that the offspring in an event can be ordered $(v_1, v_2, \dots)$ with $v_1 = v^*$. Therefore, as the parent is thought of as moving to v^* we will refer to $(v_i)_{i\geq 2}$ as the 'children' after an event. Recall that within the period of time $[0,t]$ with probability $\mathcal{O}(t^2)$ we will see two or more events. Therefore

$$\begin{aligned}
&\frac{\mathbb{E}\Big[\sum_{l(t)\in\eta_t^r}\Big(l(t)-\hat{l}(t)\Big)h(l(t)) - \sum_{l(0)\in\eta_0^r}\Big(l(0)-\hat{l}(0)\Big)h(l(0))\Big]}{t}\\
&\quad= N\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}\Bigg\{\sum_{l(0)\in\eta_t^r}\Bigg\{\Big[\Big(\mathcal{J}(l(0))-\mathcal{J}(\hat{l}(0))\Big)-\Big(l(0)-\hat{l}(0)\Big)\Big]h(l(0))\\
&\qquad\qquad+\Big(\mathcal{J}(\hat{l}(0))-\hat{l}(t)\Big)h(l(0))\\
&\quad+\Big(l(t)-\hat{l}(0)\Big)(\mathcal{J}(l(0))-l(0))\mathcal{O}(||h'||_\infty)\Bigg\}+\sum_{i=2}^{\infty}(v_i-\hat{v}_i)h(v_i)\mathbf{1}_{\{v_i \text{ is rare}\}}\Bigg\}\mathrm{d}v^*+\mathcal{O}(t),
\end{aligned}$$

where we are using $\hat{v}_i$ to denote the highest level below the ith 'child', v_i. We now require that $h(l)=0$ for all $l\geq\lambda_h$. We note that in the proof of Proposition 4.31 we have shown that the parts involving $\mathcal{J}(l(0))-l(0)$ converge in L^2 to $al(0)^2$ and so we can see that this is approximated by

$$\sum_{l(0)\in\eta_t^r}\Big\{\Big(al^2(0)-a\hat{l}^2(0)\Big)h(l(0))+\Big(l(0)-\hat{l}(0)\Big)al^2(0)\mathcal{O}(||h'||_\infty)\Big\}$$

$$+N\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}\left\{\sum_{l(0)\in\eta_t^r}\Big(\mathcal{J}(\hat{l}(0))-\hat{l}(t)\Big)h(l(0))+\sum_{i=2}^{\infty}(v_i-\hat{v}_i)h(v_i)\mathbf{1}_{\{v_i \text{ is rare}\}}\right\}\mathrm{d}v^*.$$

We note that since the ordering of the $l(t)\in\eta_t$ which are not the parent is unaffected by an event, the only possibilities that will force $\mathcal{J}(\hat{l}(0))<\hat{l}(t)$ is when a "child" has been born between the new locations of levels or when $\hat{l}(0)$ is chosen as the parent. However we use the last sum appearing in the previous equation and the fact that h is decreasing to see that the generator applied to β_t^N, for large enough N, is bounded below by

$$\begin{aligned}-C_h\beta_t^N+\sum_{l_i\in\eta_t^r}\int_{l_{i-2}}^{\frac{l_{i-2}}{1-\frac{u}{J}}}\frac{uK}{J}e^{-\frac{uK}{J}v^*}\left(v^*-\frac{l_{i-2}}{1-\frac{u}{J}}\right)\mathrm{d}v^*\\ \geq -C_h\beta_t^N-\sum_{l_i\in\eta_t^r}\mathcal{O}\Big(\frac{NK}{J^3}l_{i-2}^2\Big),\end{aligned}$$

where $C_h=\lambda_h||h'||_\infty$. We therefore see that

$$\beta_t^N+\int_0^t C_h\beta_s^N+C\frac{NKZ}{J^3}\mathrm{d}s,$$

is a sub-martingale. This concludes the proof of the lemma. □

To complete the proof of Proposition 4.32 we see that, by using Lemma 4.35,

$$\mathbb{E}\left[\left(\beta_T^N+\int_0^T C_h\beta_s^N+C\frac{NKZ}{J^3}\mathrm{d}s\right)^2\right],$$

is bounded. Therefore, by Jensen's inequality,

$$\mathbb{E}[\sup_{t\leq T}\beta_t^N]\leq\sqrt{\mathbb{E}\left[\left(\sup_{t\leq T}\beta_t^N\right)^2\right]}$$

which is also bounded by Doob's martingale inequality. Proposition 4.32 follows by application of Lemma 4.34.

4.4.3.3 Selective model - proof of Theorem 4.21

The generator of the model with selection is of the form

$$A^N f(\eta) = A^N_{neu} f(\eta) + A^N_{sel} f(\eta).$$

The analysis for the neutral part of the generator, A^N_{neu}, is the same as in Section 4.4.3.2, with the exception of a slight modification of Lemma 4.36, which we discuss in Lemma 4.39. We therefore turn our attention to A^N_{sel}, which is the part of the generator describing the selection. It can be written as

$$A^N_{sel} f(\eta) = \frac{sN}{S}\Bigg(\int_0^\infty \Bigg[\frac{uK}{J} e^{-\frac{uK}{J}v^*} g(\kappa^*, v^*) e^{-\frac{uK}{J}\int_{v^*}^\infty (1-g(\kappa^*,v))\mathrm{d}v} \times \prod_{(\kappa,l)\in\eta, l\neq l^*_{sel}} g(\kappa, \mathcal{J}_{sel}(l, l^*_{sel}, v^*)) \Bigg] \mathrm{d}v^* - f(\eta) \Bigg), \quad (4.71)$$

where we recall that the movement of the levels $\mathcal{J}_{sel}$ is specified by

$$\mathcal{J}_{sel}((l,\kappa),(l^*_{sel},\kappa^*),v^*) = \begin{cases} v^* & \text{if } l = l^*_{sel}, \\ \frac{1}{1-\frac{u}{J}}\Big(l - (l^*_{sel} - v^*)\frac{\sigma(\kappa)}{\sigma(\kappa^*)}\Big) & \text{if } l \neq l^*_{sel}, l > v^*, \\ \frac{1}{1-\frac{u}{J}} l & \text{if } l < v^*. \end{cases}$$

As in Section 4.4.3.2, we transform A^N_{sel} into a more convenient form. We use the Taylor approximation (4.52) to approximate the generator by

$$A^N_{sel} = A^N_{sel,1} + A^N_{sel,2} + \mathcal{O}\Big(\frac{K}{J} A^N_{sel,2}\Big),$$

where

$$A^N_{sel,1} f(\eta) = \frac{sN}{S}\Bigg(\int_0^\infty \frac{uK}{J} e^{-\frac{uK}{J}v^*} \Big[g(\kappa^*, v^*) \prod_{(\kappa,l)\in\eta, l\neq l^*_{sel}} g(\kappa, \mathcal{J}_{sel}(l, l^*_{sel}, v^*)) - f(\eta) \Big] \mathrm{d}v^* \Bigg), \quad (4.72)$$

$$A^N_{sel,2} f(\eta) = \frac{N}{S}\Bigg(-\int_0^\infty \frac{u^2K^2}{J^2} e^{-\frac{uK}{J}v^*} \int_{v^*}^\infty (1 - g(\kappa^*, v))\mathrm{d}v \; g(\kappa^*, v^*)$$

$$\times \prod_{(\kappa,l)\in\eta, l\neq l^*_{sel}} g(\kappa, \mathcal{J}_{sel}(l, l^*_{sel}, v^*))\mathrm{d}v^*\Bigg). \quad (4.73)$$

Once again, we treat $A^N_{sel,1}$ and $A^N_{sel,2}$ separately. We order the levels $\eta = \{(\kappa_i, l_i)\}_{i\geq 1}$ present in the system by requiring that for each i, $l_i < l_{i+1}$. Arguing in the same way as for the neutral case, we may conclude that the term involving $A^N_{sel,2}$ tends to 0 as $N \to \infty$, as the requirements $NK/J^2 \to C$ combined with $1/S \to 0$ imply that $NK/J^2S \to 0$. We turn our attention to $A^N_{sel,1}$. Recalling the difference between l^*_{neu} and l^*_{sel}, we note that for large N we will very rarely see $l^*_{sel} \neq l^*_{neu}$ and so we consider

$$A^N_{sel,3}f(\eta) = \frac{sN}{S}\Bigg(\int_0^\infty \frac{uK}{J} e^{-\frac{uK}{J}v^*} \times \Big[g(\kappa^*, v^*) \prod_{(\kappa,l)\in\eta, l\neq l^*_{neu}} g(\kappa, \mathcal{J}_{sel}(l, l^*_{neu}, v^*)) - f(\eta)\Big]\mathrm{d}v^*\Bigg). \quad (4.74)$$

We will prove Proposition 4.26 by showing that $A^N_{sel,3}$ satisfies a suitable estimate and then showing that $A^N_{sel,1} - A^N_{sel,3}$ converges to 0 by virtue of Lemma 4.38.

Lemma 4.37. *Under the conditions of Theorem 4.21,*

$$\mathbb{E}\Bigg[\Bigg|\frac{sN}{S}\Bigg(\int_0^\infty \frac{uK}{J} e^{-\frac{uK}{J}v^*}\Big[g(\kappa^*, v^*) \prod_{(\kappa,l)\in\eta_t, l\neq l^*_{neu}} g(\kappa, \mathcal{J}_{sel}(l, l^*_{neu}, v^*)) - f(\eta)\Big]\mathrm{d}v^*\Bigg) - b\Bigg(f(\eta^r_t)\sum_j \frac{g'(\kappa_i, l_i)}{g(\kappa_i, l_i)}\frac{suN}{JS} l_i\Bigg)\Bigg|\Bigg] \to 0,$$

Proof Just as in Proposition 4.31, we approximate (4.74) using (4.51) by

$$\begin{aligned}
&\frac{sN}{S}f(\eta)\Bigg(\sum_i \int_{l_{i-1}}^{l_i} \frac{uK}{J} e^{-\frac{uK}{J}v^*}\Bigg[\frac{g'(\kappa_i, l_i)}{g(\kappa_i, l_i)}(v^* - l_i) \\
&\qquad + \sum_{l\neq l_i} \frac{g'(\kappa, l)}{g(\kappa, l)}\Bigg(\frac{l\frac{u}{J}}{1-\frac{u}{J}} - \mathbf{1}_{l>l_i}\frac{\sigma(\kappa)}{\sigma(\kappa^*)}\frac{l_i - v^*}{1-\frac{u}{J}}\Bigg)\Bigg]\mathrm{d}v^*\Bigg) \\
&= Nf(\eta)\sum_i \Bigg[\frac{g'(\kappa_i, l_i)}{g(\kappa_i, l_i)}\Big(-(l_i - l_{i-1})e^{-\frac{uK}{J}l_{i-1}} + \frac{J}{uK}\Big(e^{-\frac{uK}{J}l_{i-1}} - e^{-\frac{uK}{J}l_i}\Big)\Big) \\
&\qquad + \sum_{j\neq i}\Bigg(\frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)}\frac{l_j\frac{u}{J}}{1-\frac{u}{J}}\Big(e^{-\frac{uK}{J}l_{i-1}} - e^{-\frac{uK}{J}l_i}\Big)\Bigg) \\
&+ \sum_{j\neq i}\sum_{j<i}\frac{\sigma(\kappa_j)}{\sigma(\kappa_i)}\frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)}\frac{1}{1-\frac{u}{J}}\Big(-(l_i - l_{i-1})e^{-\frac{uK}{J}l_{i-1}} + \frac{J}{uK}\Big(e^{-\frac{uK}{J}l_{i-1}} - e^{-\frac{uK}{J}l_i}\Big)\Big)\Bigg]
\end{aligned}$$

$$=: \frac{\mathcal{S}_1}{S} + \frac{sN}{S} f(\eta) \sum_i \sum_{j<i} \Big(\frac{\sigma(\kappa_j)}{\sigma(\kappa_i)} - 1 \Big) \frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)} \frac{1}{1 - \frac{u}{J}} \Big(-(l_i - l_{i-1}) e^{-\frac{uK}{J} l_{i-1}} + \frac{J}{uK} \Big(e^{-\frac{uK}{J} l_{i-1}} - e^{-\frac{uK}{J} l_i} \Big) \Big),$$

where we have used integration by parts and $\mathcal{S}_1$ is equal to the terms appearing in (4.55). Now we notice that $\mathcal{S}_1$ is multiplied by $1/S$ and since $S \to \infty$ it can be neglected. Consider then the second term. We can neglect the factor $1 - u/J$ at the cost of an error of order NK/J^3, which tends to zero. We change the order of summation to approximate it by

$$\frac{sN}{S} f(\eta) \Big(\frac{\sigma(\kappa_r)}{\sigma(\kappa_c)} - 1 \Big) \sum_j \frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)} \sum_{i<j} \Big(- (l_i - l_{i-1}) e^{-\frac{uK}{J} l_{i-1}} + \frac{J}{uK} \Big(e^{-\frac{uK}{J} l_{i-1}} - e^{-\frac{uK}{J} l_i} \Big) \Big).$$

Now we proceed precisely as in the proof of Proposition 4.31. The final approximation of $A^N_{sel,3}$ then takes the form

$$f(\eta) \Big(\frac{\sigma(\kappa_r)}{\sigma(\kappa_c)} - 1 \Big) \sum_j \frac{g'(\kappa_j, l_j)}{g(\kappa_j, l_j)} \Big(\frac{suN}{JS} l_j + \mathcal{O}\Big(\frac{1}{S} \Big) \Big).$$

□

We now turn our attention to $A^N_{sel,1} - A^N_{sel,3}$. We must treat the cases when the rare type is favoured or unfavoured separately. However, we show in both cases that $\mathbb{E}[\big| A^N_{sel,1} - A^N_{sel,3} \big|] \to 0$ and conclude with Jensen's inequality.

Lemma 4.38. *Let l^*_{neu} be the lowest level above v^*. Let l^*_{sel} be the level above v^* which minimises $\frac{l - v^*}{\sigma}$. Then*

$$\mathbb{E}\Big[\Big| \frac{sN}{S} f(\eta) \int_0^\infty \frac{uK}{J} e^{-\frac{uK}{J} v^*} g(v^*, \kappa^*) \sum_l \frac{g'(l)}{g(l)} (\mathcal{J}_{sel}(l, l^*_{sel}, v^*) - \mathcal{J}_{sel}(l, l^*_{neu}, v^*)) \mathrm{d}v^* \Big| \Big] \to 0.$$

Proof We note that for $l < v^*$, $\mathcal{J}_{sel}(l, l^*_{sel}, v^*) = \mathcal{J}_{sel}(l, l^*_{neu}, v^*)$ as

$$\mathcal{J}_{sel}(l, l^*, v^*) = \begin{cases} \frac{1}{1 - \frac{u}{J}} \Big(l - (l^* - v^*) \frac{\sigma(\kappa)}{\sigma(\kappa^*)} \Big) & l > v^*,\ l \neq l^*, \\ v^* & l = l^*, \\ \frac{1}{1 - \frac{u}{J}} l & l < v^*, \end{cases}$$

and so we need only consider

$$\frac{sN}{S}f(\eta)\int_0^\infty \frac{uK}{J}e^{-\frac{uK}{J}v^*}g(v^*,\kappa^*)\sum_{l>v^*}\frac{g'(l)}{g(l)}(l^*_{neu}-l^*_{sel})\frac{\sigma(\kappa)}{\sigma(\kappa^*)}\mathrm{d}v^*. \tag{4.75}$$

We proceed as in Section 4.4.3.2 and consider v^* in the interval $[l_{i-1}, l_i]$. We note that $l^*_{neu} = l_i$ and if l_i is of the favoured type then $l^*_{sel} = l_i$ also. If l_i is unfavoured we denote the lowest favoured type above l_i by l_i^s then $l^*_{sel} \neq l_i$ if and only if

$$l_{i-1} \leq v^* \leq \left(l_i - \frac{\sigma_w}{\sigma_s - \sigma_w}(l_i^s - l_i)\right) \wedge l_{i-1},$$

where $\sigma_w = \sigma(\kappa_w)$, $\sigma_s = \sigma(\kappa_s)$, κ_w is the unfavoured type and κ_s is the favoured type. Therefore we see that (4.75) can be written as

$$\frac{sN}{S}f(\eta)\sum_{(l_i,\kappa_i)\in\eta,\kappa_i=\kappa_w}\int_{l_{i-1}}^{(l_i-\frac{\sigma_w}{\sigma_s-\sigma_w}(l_i^s-l_i))\wedge l_{i-1}}\frac{uK}{J}e^{-\frac{uK}{J}v^*}g(v^*,\kappa^*)\sum_{l>v^*}\frac{g'(l)}{g(l)}(l^*_n-l^*_s)\frac{\sigma(\kappa)}{\sigma(\kappa^*)}\mathrm{d}v^*.$$

We can then integrate and use a Taylor expansion to see that we need only show that the following expression converges to zero:

$$\mathbb{E}\left[\frac{sN}{S}f(\eta)\sum_j\frac{g'(l_j)}{g(l_j)}\sum_{(l_i,\kappa_i)\in\eta,l_i<l_j,\kappa_i=\kappa_w}\frac{uK}{J}(l_i^s-l_i)\left(l_i-l_{i-1}-\frac{\sigma_w}{\sigma_s-\sigma_w}(l_i^s-l_i)\right)\wedge 0\right].$$

If the rare type is unfavoured we can then conclude by noting that the order of the expectation is $\mathbb{E}\Big[\frac{suNK}{SJ}(l_i^s-l_i)(l_i-l_{i-1})\Big] = \mathcal{O}\Big(\frac{NK}{SJ}\frac{1}{(K-Z)^2}\Big)$.

If the rare type is favoured we may bound it by

$$\sum_{\hat{l}_{k-1}<l_i<\hat{l}_k}\frac{suNK}{JS}(l_i-l_{i-1})^2\mathbf{1}_{\{l_i-l_{i-1}-\frac{\sigma_w}{\sigma_s-\sigma_w}(\hat{l}_k-l_i)>0\}}, \tag{4.76}$$

where we recall that $\hat{l}_k$ is the kth highest rare level. As before, we condition on there being n individuals of the common type with levels between $\hat{l}_k$ and $\hat{l}_{k-1}$. We can then consider the n levels as independent uniformly distributed rather than the ordered levels l_i. Suppose that $z, z_1, \ldots, z_{n-1}$ are independent uniformly distributed random variables on $[0,1]$, $Y = 1-z$ and $X = \min(\{z\}, \{(z-z_i)\}_{z_i<z})$. We then see that (4.76)

is bounded by

$$n\frac{suKN}{JS}\mathbb{E}\Big[X^2\mathbf{1}_{X>cY}\Big].$$

We now look at

$$\begin{aligned}
\mathbb{E}[X^2\mathbf{1}_{X>cY}] &= \int_0^1 \mathbb{E}[X^2\mathbf{1}_{X>c(1-u)}|z=u]\mathrm{d}u && (4.77)\\
&= \int_{\frac{c}{1+c}}^1 \mathbb{E}[X^2\mathbf{1}_{X>c(1-u)}|z=u]\mathrm{d}u && (4.78)\\
&= \int_{\frac{c}{1+c}}^1 \int_0^u x^2\mathbf{1}_{x>c(1-u)}\mathrm{d}\mathbb{P}[X=x|z=u]\mathrm{d}u && (4.79)\\
&= \int_{\frac{c}{1+c}}^1 \int_{c(1-u)}^u x^2\mathrm{d}\mathbb{P}[X=x|z=u]\mathrm{d}u. && (4.80)
\end{aligned}$$

Since

$$\begin{aligned}
\mathbb{P}[X>x|z=u] =& \mathbb{P}\big[[u-x,u]\cap\{z_i\}=\emptyset\big] = (1-x)^{n-1},\\
\mathrm{d}\mathbb{P}[X=x|z=u] =& (n-1)(1-x)^{n-2},
\end{aligned}$$

we may bound (4.77) by

$$\begin{aligned}
\mathbb{E}[X^2\mathbf{1}_{X>cY}] \leq& \int_{\frac{c}{1+c}}^1 \int_{c(1-u)}^1 x^2(n-1)(1-x)^{n-2}\mathrm{d}x\ \mathrm{d}u\\
=& \int_{\frac{c}{1+c}}^1 \int_0^{1-c+cu} (n-1)(1-v)^2v^{n-2}\mathrm{d}v\ \mathrm{d}u\\
\leq& \frac{n-1}{c}\left[\frac{v^n}{n(n-1)} - 2\frac{v^{n+1}}{n(n+1)} + \frac{v^{n+2}}{(n+1)(n+2)}\right]_0^1\\
=& \frac{6}{cn(n+1)(n+2)}.
\end{aligned}$$

We can then conclude by noting that

$$\sum_{n=1}^{\infty}\frac{1}{(n+1)(n+2)}\frac{(K-Z^N)^ne^{-(K-Z^N)}}{n!} = \mathcal{O}\Big(\frac{1}{(K-Z^N)^2} + \exp(-(K-Z^N))\Big),$$

which again leads to an error of order $\mathcal{O}\Big(\frac{NK}{SJ}\frac{1}{(K-Z)^2}\Big)$, completing the proof of the Lemma. □

Proof[Proof of Proposition 4.26] By Lemma 4.33, Proposition 4.26 follows from a combination of Lemma 4.37 and Lemma 4.38. □

Lemma 4.39. *Define β^N as in* (4.70)*. Then β^N is a dominated by a bounded submartingale.*

Remark 4.40. *Notice that the even though the statements of Lemmas 4.36 4.39 are the same, the behaviour of the process β^N is subtly different, as in Lemma 4.39 the movement of the levels is affected by selective events.*

Proof As in the proof of Lemma 4.36 (and using the notation introduced there)

$$\begin{aligned}
&\frac{\mathbb{E}\Big[\sum_{l(t)\in\eta_t^r}\Big(l(t)-\hat{l}(t)\Big)h(l(t))-\sum_{l(0)\in\eta_0^r}\Big(l(0)-\hat{l}(0)\Big)h(l(0))\Big]}{t}\\
&=N\int_0^\infty\Big(1+\frac{s}{S}\Big)\frac{uK}{J}e^{-\frac{uK}{J}v^*}\Bigg\{\sum_{l(0)\in\eta_t^r}\Bigg\{\Big[\Big(\mathcal{J}(l(0))-\mathcal{J}(\hat{l}(0))\Big)-\Big(l(0)-\hat{l}(0)\Big)\Big]h(l(0))\\
&\qquad\qquad+\Big(\mathcal{J}(\hat{l}(0))-\hat{l}(t)\Big)h(l(0))\\
&+\Big(l(t)-\hat{l}(0)\Big)(\mathcal{J}(l(0))-l(0))\mathcal{O}(||h'||_\infty)\Bigg\}+\sum_{i=2}^{\infty}(v_i-\hat{v}_i)h(v_i)\mathbf{1}_{\{v_i\text{ is rare}\}}\Bigg\}\mathrm{d}v^*+\mathcal{O}(t).
\end{aligned}$$

Considering combination of Proposition 4.31 and Proposition 4.26 we know that the parts involving $\mathcal{J}(l(0))-l(0)$ converge to $al(0)-bl^2(0)$. Since $\frac{s}{S}\to 0$ and observing Lemma 4.38, we may proceed as in the proof of Lemma 4.36 to conclude. □

4.4.3.4 Model with selection in fluctuating environment - proof of Theorem 4.18

Proof We begin by identifying the limit. Recall that the rescaled generator takes the form

$$A^N f(\eta,\zeta)=A^N_{neu}f(\eta,\zeta)+\hat{S}A^N_{sel}f(\eta,\zeta)+\hat{S}^2A_{env}f(\eta,\zeta), \tag{4.81}$$

where A^N_{neu} is defined as in (4.50), A^N_{sel} is defined as in (4.71), and

$$A_{env}f(\eta,\xi)=\mathbb{E}_\pi[f(\eta,\xi)]-f(\eta,\xi).$$

We also use (4.42) and that $A_{neu}f_1$ is of order 1 by previous calculations. To identify the correct limit it is therefore enough to evaluate $A^N_{sel}(f_1(\eta,\xi))$, which can be approximated as

$$A^N_{sel}(f_1(\eta,\xi)) = A^N_{sel}\left(f(\eta)\frac{suN}{JS}\left(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}-1\right)\sum_j l_j\frac{g'(\kappa_j,l_j)}{g(\kappa_j,l_j)}\right)$$
$$= \left[\frac{suN}{JS}\left(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}-1\right)\right]^2 f(\eta)$$
$$\times\left\{\left(\sum_{i\neq j} l_i l_j\frac{g'(\kappa_i,l_i,)g'(\kappa_j,l_j)}{g(\kappa_i,l_i)g(\kappa_j,l_j)}\right)+\left(\sum_j l_j\frac{g'(\kappa_j,l_j)+l_j g''(\kappa_j,l_j)}{g(\kappa_j,l_j)}\right)\right\}+\mathcal{O}\left(\frac{1}{S}+\frac{1}{K^{1/2}}\right)$$

The bound (3.22) follows directly from proofs of Propositions 4.26, 4.31, 4.32, with Lemma 4.36 modified in a way analogous to Lemma 4.39. □

4.5 Scaling limits of the SLFV - dynamics of the rare type

In this section we are interested in a spatial analogue of the results from Section 4.4. We show that under a certain scaling, the dynamics of the subpopulation with a rare mutation, which is a part of a population evolving according to a version of the Spatial-Lambda-Fleming-Viot model with selection in a fluctuating environment, is given by a superBrownian motion in a random environment. Since under our scaling the proof of the result does not differ significantly from the one discussed in Section 4.4, our discussion will be rather brief and focus on highlighting the differences and required modifications.

We begin with a description of the model. As before, we consider a population with two genetic types, rare and common, which we denote by κ_r and κ_c, respectively. We also consider the random field ζ, specified by Definition 4.2. Since in our model we consider a population with a countable number of individuals, the state of the population can be represented as

$$\eta = \sum_{(x,\kappa,l)} \delta_{x,\kappa,l}.$$

We assume that η is a conditionally Poisson system with Cox measure $\Xi \times m_{leb}$.

The evolution of the population is determined by reproduction events of two types - neutral and selective driven by independent Poisson point processes Π^{neu} and Π^{sel}, which specify the time, location, impact and radius of the events. They are analogous to those in Section 4.4, but now we assume that only individuals within the ball centred at the location of the event, of radius specified during event, are affected. A rigorous definition of the model follows.

Definition 4.41 (Lookdown representation of SLFVSRE)**.** *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 $r \mapsto \nu_r$ is measurable and*

$$\int_{(0,\infty)} r^d \int_{[0,1]} u\, \nu_r(\mathrm{d}u)\mu(\mathrm{d}r) < \infty. \tag{4.82}$$

Fix $s \in (0,1)$. Let Π^{neu}, Π^{sel} be a pair of independent Poisson processes with intensity measures $(1-s)\mathrm{d}t \otimes \mathrm{d}y \otimes \mu(\mathrm{d}r)\nu_r(\mathrm{d}u)$ and $s\mathrm{d}t \otimes \mathrm{d}y \otimes \mu(\mathrm{d}r)\nu_r(\mathrm{d}u)$ respectively.

Let Π^{env} be a Poisson process independent of Π^{neu}, Π^{sel}. The lookdown representation of SLFVSRE is a process taking values in purely atomic measures on $\{-1,1\} \times \mathbb{R}^d \times \mathbb{R} \times \{\kappa_r, \kappa_c\}$ with dynamics described as follows.

1. *If $(t,y,r,u) \in \Pi^{neu}$*

   (a) *a group of new individuals with levels $(v_1, v_2, \dots,)$ is added to the population within the ball $B_r(y)$. Their levels are distributed according to a Poisson process with intensity u.*

   (b) *Let $v^* = \min\{v_1, v_2, \dots,\}$. The type of the new individuals is chosen to be the same as the type of an individual with the lowest level above v^* within the ball $B_r(y)$.*

   (c) *As a result of an event the levels are changed according to*

$$\mathcal{J}_{neu}(l, l^*, v^*, x, (y,r)) = \begin{cases} l & \text{if } x \notin B_r(y), \\ \frac{1}{1-\frac{u}{J}}(l - (l^* - v^*)) & \text{if } l > l^*, x \in B_r(y), \\ \frac{1}{1-\frac{u}{J}} l & \text{if } l < l^*, x \in B_r(x), \\ v^* & \text{if } l = l^*, x \in B_r(y). \end{cases}$$

2. If $(t, y, r, u) \in \Pi^{sel}$

(a) a group of new individuals with levels $(v_1, v_2, \dots,)$ *is added to the population. Their levels distributed according to a Poisson process with intensity* u*.*

(b) Let $v^* = \min\{v_1, v_2, \dots, \}$*. The type of the new individuals is chosen to be the same as the type of an individual with level above* $v*$ *minimizing* $(l_i - v^*)/\sigma(\kappa_i, \zeta)$ *within the ball* $B(x, r)$*.*

(c) As a result of an event the levels are changed according to

$$\mathcal{J}_{sel}(l, l^*, v^*, x, (y, r)) = \begin{cases} l & \text{if } x \notin B_r(y), \\ \frac{1}{1-\frac{u}{J}}(l - (l^* - v^*)\frac{\sigma(\kappa,\zeta)}{\sigma(\kappa^*,\zeta)}) & \text{if } l > l^*, x \in B_{r/M}(y), \\ \frac{1}{1-\frac{u}{J}} l & \text{if } l < l^*, x \in B_r(y), \\ v^* & \text{if } l = l^*, x \in B_r(y). \end{cases}$$

3. The dynamics of Π^{env} *are specified by Definition 4.2.*

Definition 4.41 is more general than we require, however, we include it to underline the possibility of extending our results. However, in the interest of keeping our notation as simple as possible, from now on we shall specialise to fix the radius and impact of reproduction events.

Assumption 4.42. *From now on, fix* $R \in (0, \infty)$ *and* $\bar{u} \in (0, 1)$ *and take*

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

The integrability condition (4.82) is trivially satisfied for the model with fixed radius and impact.

4.5.1 Scaling and statement of main results

At the Nth stage of our scaling, the local population density will be $K = K(N)$. We shall denote the representative of the density of the scaled SLFVSRE by w^N and the population of rare individuals by $X^N = Kw^N$, which is defined Lebesgue almost

everywhere. We shall think of X^N as a measure-valued process and abuse notation by writing, for any Borel measurable ϕ,

$$\left\langle X_t^N, \phi \right\rangle = K \int_{\mathbb{R}^d} \phi(x) w_t^N(x) \mathrm{d}x = \int_{\mathbb{R}^d} \phi(x) X_t^N(x) \mathrm{d}x.$$

The scaling is nearly the same as in Chetwynd-Diggle and Etheridge (2018). For the scaled process, time is sped up by a factor N, space is shrunk by $M(N)$, and the impact of each event is reduced by a factor $J(N)$. The rate of the selective events is multiplied by $\widehat{S}/S(N)$. Let $C(d) := \int_{\mathcal{B}_1(0)} x^2 \mathrm{d}x$.

Theorem 4.43. *Suppose that X_0^N is absolutely continuous with respect to Lebesgue measure, that the support* $\mathrm{supp}(X_0^N) \subseteq D$*, where D is a compact subset of $\mathbb{R}^d$ (independent of N), and that X_0^N converges weakly to X_0. Moreover, suppose that, as N tends to infinity,*

$$\frac{C_d u r^{d+2} N}{JM^2} \to C_1;\ J, K, M, S, \widehat{S} \to \infty;\ \frac{K}{JM^d} \to 0;\ \frac{u^2 V_R N K}{J^2 M^d} \to a;$$
$$\frac{N^2}{KJ^2M^d} \to 0;\ \mathbb{E}_\pi\left[\left\{\frac{suNV_R}{SJ}\left(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)} - 1\right)\right\}^2\right] \to b^2;\ \frac{\widehat{S}}{S} \to 0.$$

In addition, assume that there exists an m such that, as N tends to infinity, $N(K/J)^m \to 0$. Then the sequence $X_N(t)$ converges weakly to superBrownian motion in a random environment with initial condition X_0, diffusion parameter C_1, drift b^2, quadratic variation parameter $2a$.

As in Section 4.4 we record two theorems which are a by-product of our technique.

Theorem 4.44. *Suppose that X_0^N is absolutely continuous with respect to Lebesgue measure, that the support* $\mathrm{supp}(X_0^N) \subseteq D$*, where D is a compact subset of $\mathbb{R}^d$ (independent of N), and that X_0^N converges weakly to X_0. Moreover, suppose that, as N tends to infinity,*

$$\frac{C_d u r^{d+2} N}{JM^2} \to C_1; \quad J, K, M \to \infty; \quad \frac{K}{JM^d} \to 0; \quad \frac{N^2}{M^d K J^2} \to 0; \quad \frac{u^2 V_R N K}{J^2 M^d} \to a,$$

In addition, assume that there exists an n such that, as N tends to infinity, $N(K/J)^n \to 0$. Then the sequence $X_N(t)$ converges weakly to superBrownian motion with initial condition X_0 diffusion parameter C_1, quadratic variation parameter $2a + q(x, y)$.

Theorem 4.45. *Suppose that X_0^N is absolutely continuous with respect to Lebesgue measure, that the support $\mathrm{supp}(X_0^N) \subseteq D$, where D is a compact subset of $\mathbb{R}^d$ (independent of N), and that X_0^N converges weakly to X_0. Moreover, suppose that, as N tends to infinity,*

$$\frac{N}{JM^2} \to C_1 \quad J, K, M \to \infty; \quad \frac{K}{JM^d} \to 0; \quad \frac{N^2}{M^d K J^2} \to 0; \quad \frac{u^2 V_R N K}{J^2 M^d} \to a$$
$$\frac{suNV_R}{JS}\left(\frac{\sigma(\kappa_r, \zeta)}{\sigma(\kappa_c, \zeta)} - 1\right) \to b,$$

In addition, assume that there exists an n such that, as N tends to infinity, $N(K/J)^n \to 0$. Then the sequence $X_N(t)$ converges weakly to a critical superBrownian motion with initial condition X_0, diffusion parameter C_1, drift b, and quadratic variation parameter $2a$.

Remark 4.46. *We note that the 'sparsity' condition we require for the neutral case is stronger than the one in Chetwynd-Diggle and Etheridge (2018). Our set of conditions implies $M^2/J \to 0$ in comparison to theirs*

$$\frac{M}{J} \to 0 \text{ if } d = 1, \quad \frac{\log M}{J} \to 0 \text{ if } d = 2, \quad \frac{1}{J} \to 0 \text{ if } d \geq 3.$$

It is due to the fact that the lookdown construction 'sees' the Hausdorff dimension of the support of superBrownian motion in a way previous work did not, since our set of test functions do not smooth out the support the process, in sharp contrast to their approach. We require the intensity of levels within the ball of radius r/M to tend to infinity in order to allow the limit to have Hausdorff dimension two. This observation explain why the proof of our results in the spatial case does not differ significantly from the proof in the non-spatial case.

The strategy of the proof is nearly entirely analogous to that laid out in Subsection 4.4.3. We therefore focus on describing the differences. In a sharp contrast to situation described in the Remark 4.27 we do not claim to show any results on the

limits of sequences of lookdown representations. We focus on the result for the projected version of the model. We divide the generator of the lookdown representation into two separate parts - one describes the spatial movement of the particles, the other describes the evolution of the levels. We show that the selective events (and therefore the fluctuations in the direction of selection) do not affect the movement of the particles in the limit. For the convergence of the neutral model we refer to results of Chetwynd-Diggle and Etheridge (2018). We then use the lookdown representation to deduce the right form of the limiting generator and to justify the separation of timescales trick for the selective part of the generator. The part of the proof which deals with the evolution of the levels is analogous to that of Section 4.4, and we do not repeat it here. For the reader's convenience we include intensity estimate and discuss the splitting of the generator.

Once again we consider a stopped process, see Remark 4.28. The domain of the generator is specified by (4.24), (4.25).

4.5.2 Intensity estimate

In complete analogy to the non-spatial version of the model we observe that the generator of the projected process is given by

$$\begin{aligned}
\mathcal{L}^N f(\langle \phi, X_0^N\rangle) = NM^d\Bigg[&\int_{\mathbb{R}^d}\int_{B_r^M(x)} \frac{1}{|B_r^M|} w_0^N(z) f\Bigg(K\frac{u}{J}\int_{B_r^M(x)}\phi(y)\mathrm{d}y \\
&+ K\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Bigg) - f\Big(K\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\mathrm{d}x \\
&+ \int_{\mathbb{R}^d}\int_{B_r^M(x)} (1-w_0^N(z)) f\Bigg(K\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Bigg) \\
&- f\Big(K\int_{\mathbb{R}^d}\phi(y)w_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\ \mathrm{d}x\Bigg] \\
+N\frac{s\widehat{S}}{S}M^d\Bigg[&\int_{\mathbb{R}^d}\int_{B_r^M(x)} \frac{1}{|B_r^M|}\frac{\sigma(\kappa_r,\zeta)w_0^N(z)}{\sigma(\kappa_r,\zeta)w_0^N(z)+\sigma(\kappa_s,\zeta)(1-w_0^N(z))} f\Bigg(K\frac{u}{J}\int_{B_r^M(x)}\phi(y)\mathrm{d}y \\
&+ K\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Bigg) - f\Big(K\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\Big)\mathrm{d}z\mathrm{d}x
\end{aligned}$$

$$+\int_{\mathbb{R}^d}\int_{B_r^M(x)}\frac{\sigma(\kappa_c,\zeta)(1-w_0^N(z))}{\sigma(\kappa_r,\zeta)w_0^N(z)+\sigma(\kappa_s,\zeta)(1-w_0^N(z))}f\bigg(K\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)w_0^N(y)\mathrm{d}y\bigg)$$
$$-f\bigg(K\int_{\mathbb{R}^d}\phi(y)w_0^N(y)\mathrm{d}y\bigg)\mathrm{d}z\,\mathrm{d}x\bigg]. \quad (4.83)$$

The first two terms represent the part of the generator describing the effect of the neutral events and is the same as in Chetwynd-Diggle and Etheridge (2018). The last two terms represent the effects of selective events. Substituting $Kw_0^N = X_0^N$ this becomes

$$\begin{aligned}
\mathcal{L}^N f(\langle\phi, X_0^N\rangle) &= \mathcal{L}^N_{neu} f(\langle\phi, X_0^N\rangle) + \mathcal{L}^N_{sel} f(\langle\phi, X_0^N\rangle)\\
&NM^d\bigg[\int_{\mathbb{R}^d}\int_{B_r^M(x)}\frac{1}{|B_r^M|}\frac{X_0^N(z)}{K}f\bigg(K\frac{u}{J}\int_{B_r^M(x)}\phi(y)\mathrm{d}y\\
&+\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\bigg)-f\bigg(\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\bigg)\mathrm{d}z\mathrm{d}x\\
&+\int_{\mathbb{R}^d}\int_{B_r^M(x)}\Big(1-\frac{X_0^N(z)}{K}\Big)f\bigg(\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\bigg)\\
&-f\bigg(\int_{\mathbb{R}^d}\phi(y)X_0^N(y)\mathrm{d}y\bigg)\mathrm{d}z\,\mathrm{d}x\bigg]\\
+N\frac{s\widehat{S}}{S}M^d&\bigg[\int_{\mathbb{R}^d}\int_{B_r^M(x)}\frac{1}{|B_r^M|}\frac{\sigma(\kappa_r,\zeta)\frac{X_0^N(z)}{K}}{\sigma(\kappa_r,\zeta)\frac{X_0^N(z)}{K}+\sigma(\kappa_s,\zeta)\Big(1-\frac{X_0^N(z)}{K}\Big)}f\bigg(K\frac{u}{J}\int_{B_r^M(x)}\phi(y)\mathrm{d}y\\
&+\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\bigg)-f\bigg(\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\bigg)\mathrm{d}z\mathrm{d}x\\
&+\int_{\mathbb{R}^d}\int_{B_r^M(x)}\frac{\sigma(\kappa_c,\zeta)\Big(1-\frac{X_0^N(z)}{K}\Big)}{\sigma(\kappa_r,\zeta)\frac{X_0^N(z)}{K}+\sigma(\kappa_s,\zeta)\Big(1-\frac{X_0^N(z)}{K}\Big)}f\bigg(\Big(1-\frac{u}{J}\Big)\int_{B_r^M(x)}\phi(y)X_0^N(y)\mathrm{d}y\bigg)\\
&-f\bigg(\int_{\mathbb{R}^d}\phi(y)X_0^N(y)\mathrm{d}y\bigg)\mathrm{d}z\,\mathrm{d}x\bigg]. \quad (4.84)
\end{aligned}$$

We state useful lemmas from Chetwynd-Diggle and Etheridge (2018), which describe the form and estimates for the neutral part of the generator. We need some notation.

Definition 4.47. *We denote by $\mathcal{A}^N$ the operator*

$$\mathcal{A}^N(\phi) := \frac{C(d)Nur^{d+2}}{JM^2}\Delta\phi,$$

with $C(d) := \int_{\mathcal{B}_1(0)} x^2\mathrm{d}x$.

Lemma 4.48 (Chetwynd-Diggle and Etheridge (2018), Lemma 4.2)**.** *For $f(x,\zeta)=x$ and $\phi_s(x):\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}\in C_0^{2,3}$,*

$$\int_0^t \mathcal{L}^N_{neu}(\langle\phi,X^N(s)\rangle)\mathrm{d}s=\int_0^t\Big\langle X^N_s,\dot{\phi}_s\Big\rangle+\Big\langle X^N_s,\mathcal{A}^N(\phi_s)\Big\rangle\mathrm{d}s+N^N_t(\phi), \tag{4.85}$$

where

$$|N^N_t(\phi)|\leq\mathcal{O}\bigg(\frac{N\sup_{0\leq s\leq t}\|\phi_s\|_{C^3}}{JM^3}V_R\bigg)\int_0^t\Big\langle X^N_s,\mathbf{1}\Big\rangle\mathrm{d}s. \tag{4.86}$$

We proceed to the proof of Lemma 4.49. We notice that the proof is very reminiscent of our proof of Lemma 4.24 and the proof of Chetwynd-Diggle and Etheridge (2018), Lemma 5.6.

Lemma 4.49. *Let $X=Kw$ denote the total intensity of individuals of the rare type. Assume that $\mathbb{E}[X^N(0)]<\infty$. Then for any $T>0$*

$$\sup_{t\leq T}\sup_N\mathbb{E}[\langle X^N,1\rangle]<\infty, \tag{4.87}$$

$$\lim_{H\to\infty}\sup_N\mathbb{P}\bigg[\sup_{t\leq T}\langle X^N,1\rangle\geq H\bigg]=0. \tag{4.88}$$

Proof We observe that with the test functions chosen as in Lemma 4.48, the part of the generator describing the change in the population resulting from a selective event can be written as

$$\begin{aligned}
&\mathcal{L}^N_{sel}(\langle\phi,X^N(s)\rangle)=\\
&N\frac{us\widehat{S}}{SJ}M^d\bigg[\int_{\mathbb{R}^d}\fint_{B^M_r(x)}\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}\frac{X(z)}{K}\int_{B^M_r(x)}(K-X(y))\phi(y)\mathrm{d}y\mathrm{d}z\mathrm{d}x\\
&\quad-\int_{\mathbb{R}^d}\fint_{B^M_r(x)}\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}\Big(1-\frac{X(z)}{K}\Big)\int_{B^M_r(x)}X(y)\phi(y)\mathrm{d}y\mathrm{d}z\mathrm{d}x\bigg]\\
&\quad=N\frac{us\widehat{S}}{SJ}M^d\bigg[\int_{\mathbb{R}^d}\fint_{B^M_r(x)}\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_r,\zeta)\frac{X}{K}+\sigma(\kappa_c,\zeta)\Big(1-\frac{X}{K}\Big)}X(z)\int_{B^M_r(x)}\phi(y)\mathrm{d}y\mathrm{d}z\mathrm{d}x\\
&\qquad\qquad-\int_{\mathbb{R}^d}\fint_{B^M_r(x)}\int_{B^M_r(x)}X(y)\phi(y)\mathrm{d}y\mathrm{d}z\mathrm{d}x\bigg].
\end{aligned} \tag{4.89}$$

We observe that by Taylor's Theorem $\phi(y)$ can be locally approximated by

$$\phi(y) = \phi(z) + \nabla\phi(z)(y-z) + \|\phi\|_{C^2(\mathbb{R}^d)}\mathcal{O}\Big(|y-z|^2\Big).$$

Therefore by using a calculation analogous to (4.47) and the fact that $|y-z| < R/M$ within a ball $B_r(x)$ we may approximate (4.89) by

$$\begin{aligned}&\mathcal{L}^N_{sel}(\langle\phi, X^N(s)\rangle) = \\ &N\frac{us\widehat{S}}{SJM^{2d}}\left(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}-1\right)\langle X,\phi\rangle + \frac{1}{K}\langle X,\phi\rangle^2\mathcal{O}\left(C_\sigma N\frac{us\widehat{S}}{SJM^{2d}}\right) + \mathcal{O}\left(\frac{N\widehat{S}}{SJM}\right), \quad (4.90)\end{aligned}$$

where C_σ is a constant depending on σ.

Let $\chi^N(t) = \mathbb{E}[\langle X^N_t, 1\rangle]$. Let h_R denote a sequence of smooth functions such that h_R is supported on the ball of radius $2R$ centred at zero and is equal to 1 on the ball of radius R centred at zero.

Assume in addition that the sequence h_R satisfies

$$\Delta h_R \leq \epsilon; \quad h_R \leq h_{R+1}.$$

We combine (4.85) and (4.90) and take expectation in (4.84) to obtain

$$\begin{aligned}\mathbb{E}\Big[\langle X^N_t,\phi\rangle\Big] &= \mathbb{E}\Big[\langle X^N_0,\phi\rangle\Big] + \mathbb{E}\left[\int_0^t\Big\langle X^N_s,\dot{\phi}_s\Big\rangle + \Big\langle X^N_s,\mathcal{A}^N(\phi_s)\Big\rangle\mathrm{d}s\right] \\ &+ N\frac{us\widehat{S}}{SJM^{2d}}\mathbb{E}\left[\left(\frac{\sigma(\kappa_r,\zeta)}{\sigma(\kappa_c,\zeta)}-1\right)\int_0^t\langle X,\phi\rangle + \frac{C_\sigma}{K}\langle X,\phi\rangle^2\mathrm{d}s\right] + \mathbb{E}\Big[N^N_t(\phi)\Big] \\ &\leq \chi^N(0) + C\|\Delta h_R\| + \mathcal{O}\left(\frac{N\sup_{0\leq s\leq t}\|\phi_s\|_{C^3}}{JM^3}V_R + N\frac{us\widehat{S}C_\sigma}{SJM^{2d}}\right)\chi^N(0), \quad (4.91)\end{aligned}$$

where we have used the properties of h_R. Letting R tend to infinity and using the Monotone Convergence Theorem, we arrive at

$$\chi^N(t) \leq \chi^N(0) + C\|\Delta h_R\| + \mathcal{O}\left(\frac{N\sup_{0\leq s\leq t}\|\phi_s\|_{C^3}}{JM^3}V_R + N\frac{us\widehat{S}C_\sigma}{SJM^{2d}}\right)\chi^N(0)$$

We now apply Grönwall's inequality to conclude. The second part of the statement follows exactly as the second part of the proof of Lemma 4.24. □

4.5.3 Sketch of the proof of Theorem 4.43

The generator of the SLFVRE process is given by

$$
\begin{aligned}
&A^N f(\eta) = \\
&\quad NM^d \int_{\mathbb{R}^d} \Bigg(\int_0^\infty \Bigg[\frac{uK|B_r|}{JM^d} e^{-\frac{uK|B_r|}{JM^d} v^*} \hat{g}_{y,r/M}(\kappa^*, v^*) e^{-\frac{uK|B_r|}{JM^d} \int_{v^*}^\infty (1-\hat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v} \\
&\qquad \times \prod_{(x,\kappa,l)\in\eta, l\neq l^*} g\Big(x,\kappa,\mathcal{J}_{neu}(l,l^*,v^*,x,(y,r/M))\Big) \Bigg] \mathrm{d}v^* - f(\eta) \Bigg) \mathrm{d}y \\
&+ N\frac{s\widehat{S}}{S} M^d \int_{\mathbb{R}^d} \Bigg(\int_0^\infty \Bigg[\frac{uK|B_r|}{JM^d} e^{-\frac{uK|B_r|}{JM^d} v^*} \hat{g}_{y,r/M}(\kappa^*, v^*) e^{-\frac{uK|B_r|}{JM^d} \int_{v^*}^\infty (1-\hat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v} \\
&\qquad \times \prod_{(x,\kappa,l)\in\eta, l\neq l^*} g\Big(x,\kappa,\mathcal{J}_{sel}(l,l^*,v^*,x,(y,r/M))\Big) \Bigg] \mathrm{d}v^* - f(\eta) \Bigg) \mathrm{d}y + A^{env}, \quad (4.92)
\end{aligned}
$$

where A^{env} is specified as in (4.5) and $\hat{g}_{y,r/M}(\kappa,l) := \frac{1}{|B_{r/M}|} \int_{B_{r/M}(y)} g(z,\kappa,l)\mathrm{d}z$.

We split the generator into three parts, by adding and subtracting $g(y,\kappa^*,v)$ inside the integral. The first two parts describes the movement of the levels.

$$
\begin{aligned}
&A^N_{full,neu} f(\eta) \\
&\quad = NM^d \int_{\mathbb{R}^d} \Bigg(\int_0^\infty \Bigg[\frac{uK|B_r|}{JM^d} e^{-\frac{uK|B_r|}{JM^d} v^*} g(x^*,\kappa^*,v^*) e^{-\frac{uK|B_r|}{JM^d} \int_{v^*}^\infty (1-\hat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v} \\
&\prod_{(x,\kappa,l)\in\eta_{B_{r/M}(y)}, l\neq l^*} g\Big(x^*,\kappa,\mathcal{J}_{neu}(l,l^*,v^*,x,(y,r/M))\Big) \prod_{(x,\kappa,l)\notin\eta_{B_{r/M}(y)}, l\neq l^*} g\Big(x,\kappa,l\Big) \Bigg] \mathrm{d}v^* \\
&\qquad\qquad - f(\eta) \Bigg) \mathrm{d}y, \quad (4.93)
\end{aligned}
$$

where $\eta_{B_{r/M}(y)} := \{(x,\kappa,l) \in \eta : x \in B_{r/M}(y)\}$ and

$$
\begin{aligned}
&A^N_{full,sel} f(\eta) \\
&\quad = N\frac{s\widehat{S}}{S} M^d \int_{\mathbb{R}^d} \Bigg(\int_0^\infty \Bigg[\frac{uK|B_r|}{JM^d} e^{-\frac{uK|B_r|}{JM^d} v^*} g(x^*,\kappa^*,v^*) e^{-\frac{uK|B_r|}{JM^d} \int_{v^*}^\infty (1-\hat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v} \\
&\prod_{(x,\kappa,l)\in\eta_{B_{r/M}(y)}, l\neq l^*} g\Big(x^*,\kappa,\mathcal{J}_{sel}(l,l^*,v^*,x,(y,r/M))\Big) \prod_{(x,\kappa,l)\notin\eta_{B_{r/M}(y)}, l\neq l^*} g\Big(x,\kappa,l\Big) \Bigg] \mathrm{d}v^* \\
&\qquad\qquad - f(\eta) \Bigg) \mathrm{d}y, \quad (4.94)
\end{aligned}
$$

Observe that $\widehat{g})_{y,R/M}$ is symmetrical with respect to y. Therefore (4.93), (4.94) can be treated using arguments which are analogous to those which we applied to $A_{neu,1}, A_{neu,2}, A_{sel,1}$ and $A_{sel,2}$ in Section 4.4. The second two parts are given by

$$A^N_{2,neu}f(\eta) = NM^d \int_{\mathbb{R}^d} \Bigg(\int_0^\infty \Bigg[\frac{uK|B_r|}{JM^d} e^{-\frac{uK|B_r|}{JM^d}v^*} \Big(\hat{g}_{y,r/M}(\kappa^*, v^*) - g(y, \kappa^*, v^*) \Big) \\ \times e^{-\frac{uK|B_r|}{JM^d}\int_{v^*}^\infty (1-\hat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v} \\ \times \prod_{(x,\kappa,l)\in\eta, l\neq l^*} g\Big(x, \kappa, \mathcal{J}_{neu}(l, l^*, v^*, x, (y, r/M))\Big) \Bigg] \mathrm{d}v^* \Bigg) \mathrm{d}y \quad (4.95)$$

and

$$A^N_{2,sel}f(\eta) = N\frac{s\widehat{S}}{S}M^d \int_{\mathbb{R}^d} \Bigg(\int_0^\infty \Bigg[\frac{uK|B_r|}{JM^d} e^{-\frac{uK|B_r|}{JM^d}v^*} \Big(\hat{g}_{y,r/M}(\kappa^*, v^*) - g(y, \kappa^*, v^*) \Big) \\ \times e^{-\frac{uK|B_r|}{JM^d}\int_{v^*}^\infty (1-\hat{g}_{y,r/M}(\kappa^*,v))\mathrm{d}v} \\ \times \prod_{(x,\kappa,l)\in\eta, l\neq l^*} g\Big(x, \kappa, \mathcal{J}_{sel}(l, l^*, v^*, x, (y, r/M))\Big) \Bigg] \mathrm{d}v^* \Bigg) \mathrm{d}y. \quad (4.96)$$

We observe that our scaling implies that if the contribution from (4.95) is non-negligible, the contribution from (4.96) vanishes in the limit. For that reason the results of Chetwynd-Diggle and Etheridge (2018) are sufficient to deduce the behaviour of the spatial movement of the generator of all cases of interest. From

$$A^N_2 f(\eta) = NM^d \int_{\mathbb{R}^d} \int_0^\infty \frac{uK|B_r|}{JM^d} e^{-\frac{uK|B_r|}{JM^d}v^*} \Big(\hat{g}_{y,r/M}(\kappa^*, v^*) - g(x^*, \kappa^*, v^*) \Big) \\ \times e^{-\frac{uK|B_r|}{JM^d}\int_{v^*}^\infty (1-\hat{g}_{y,r/m}(\kappa^*,v))\mathrm{d}v} \prod_{(x,\kappa,l)\in\eta, l\neq l^*} g(x, \kappa, \mathcal{J})\mathrm{d}v^*\mathrm{d}y,$$

we can see as in the calculations in Section 4.4 that this leads to

$$\alpha A^N_2 h(\Xi) \\ = NM^d \int_{\mathbb{R}^d\times\mathcal{K}} \frac{M^d}{|B_r|} \int_{B_{r/M}(x^*)} e^{-\int_{\mathbb{R}/B_{r/M}(y)} h(z,\kappa)\Xi(\mathrm{d}z,\mathrm{d}\kappa)} e^{-\left(1-\frac{u}{J}\right)\int_{B_{r/M}(y)} h(z,\kappa)\Xi(\mathrm{d}z,\mathrm{d}\kappa)} \\ \times \Big[e^{-\frac{uK|B_r|}{JM^d}\hat{h}_{y,r/M}(\kappa^*)} - e^{-\frac{uK|B_r|}{JM^d}h(x^*,\kappa^*)} \Big] \mathrm{d}y \; \Xi(\mathrm{d}x^*, \mathrm{d}\kappa^*).$$

Performing a simple Taylor expansion and noting this is identical to calculations appearing in Chetwynd-Diggle and Etheridge (2018), Section 4.3 we see that this will give us

$$\alpha A_2^N f(\Xi) = \exp(-\langle h, \Xi\rangle)\left\langle \frac{C(d)ur^dN}{JM^2}\Delta h + \mathcal{O}(1/M), \Xi \right\rangle,$$

where we recall that $C(d) := \int_{B_1} x^2 \mathrm{d}x$ and $\langle h, \Xi\rangle := \int_{\mathbb{R}^d\times\mathcal{K}} h(x,\kappa)\Xi(\mathrm{d}x, \mathrm{d}\kappa)$.

For the sake of completeness we state the propositions describing the evolution of the levels. We observe that both (4.93) and (4.94) are integrals over compact sets (recall that $g(x,l)$ is equal to 1 outside of a compact set) of their non-spatial counterparts studied in Section 4.4.3. The task of analysing this generator is then a simple expansion of arguments in that section. In particular, the sparsity condition $K/M^d \to \infty$ leads to following analogous of Proposition 4.25 and Proposition 4.26.

Proposition 4.50. *Under the conditions of Theorem 4.44,*

$$\mathbb{E}\Bigg[\sup_{t\le T}\Bigg|\int_0^t A^N_{full,neu} f(\eta_s) \\ - \left(f(\eta_s^r)\sum_{l_i(s)\in\eta_s^N} a l_i^2 \frac{\partial_l g(x_i,l_i(s))}{g(x_i,l_i(s))} + 2af(\eta_s^r)\sum_{l_i(s)\in\eta_s^r}\int_{l_i(s)}^\infty (1-g(x_i,v))\mathrm{d}v\right)\mathrm{d}s\Bigg|\Bigg] \to 0.$$

Proposition 4.51. *Under the conditions of Theorem 4.45 for any $T \in \mathbb{R}$*

$$\mathbb{E}\left[\sup_{t\le T}\left|\int_0^t A^N_{full,sel} f(\eta_s^N) - f(\eta_s^N)\left(\sum_{l_i(t)\in\eta_t^N} -bl_i\frac{g'(x_i,l_i)}{g(x_i,l_i)}\right)\mathrm{d}s\right|\right] \to 0.$$

The proof of the Theorem 4.43 now follows from application of Theorem 3.16 to the process characterised by the generator of the projected version of $A^N_{full,neu}$, $A^N_{full,sel}$ and A_{env}.

Concluding remarks

In this thesis we have examined the impact of the rapidly fluctuating selection on the evolution of a population governed by the (spatial) Lambda-Fleming-Viot model. We have identified diffusion approximations for both the non-spatial and spatial versions of the model, which allows us to model the behaviour of the genetic frequencies in the population forward in time. We have shown that in the non-spatial setting the scaling limit is an SDE in which two terms (correction to balancing selection and a random term) directly describes the influence of the fluctuations in the environment. In the spatial setting the corresponding limit is an SPDE. In dimensions greater than one, the 'genetic drift', as usual in this scaling regime, has disappeared, but some stochasticity is still present in the limit due to fluctuations in the environment.

We have also studied the behaviour of an establishing mutation within a population evolving according to the models described above. We have shown that under a suitable rescaling, it is described by the Feller diffusion in a random environment in the non-spatial setting and by the SuperBrownian motion in a random environment for a population distributed across a spatial continuum. Both of those line of investigation presented some technical challenges.

Our results open some avenues for further exploration. From the biological point of view, the most interesting one lies in the description of genealogies connected to the evolution described by our models, as those are what we are trying to infer from observable data. The most interesting question from that perspective is whether it is possible to find ways to distinguish (in data) patterns of allele frequencies driven by spatially fluctuating selection and temporally fluctuating selection. In this thesis our

only step in that direction is the introduction of the system of equations for tracer dynamics.

The usual way of identifying model for genealogies is via duality methods. The presence of fluctuations makes this question less straightforward to answer. Apart from rather limited situations, we were not able to establish a useful duality (or even approximate duality) for our models. However, if one could find an invariant measure for the SPDE describing the scaling limit, one could employ the idea of coalescence in random background, which has been introduced in the context of classical non-spatial models by Barton et al. (2004).

Description of genealogies does not become easier for the SuperBrownian motion in random environment. Even the lookdown representation does not make this question easier - for the Branching Brownian motion, the genealogy can be recovered from the locations and the values of levels only when the birth rates and death rates are constant, see Kurtz and Rodrigues (2011), Section 3.3.

There are also other interesting questions related to the models discussed in the thesis. A better understanding of more varied mechanisms of selection (for example, frequency dependent models), questions on the speed of propagation of a new mutation, possible inclusion of geographical barriers are just a few examples of questions of interest. From a more mathematical perspective, some interesting further directions include a better understanding of lookdown representations, study of qualitative and quantitative properties of limiting models and development of a more robust framework for study of processes in random environment, at least in a setting similar to the one presented in this thesis.

Bibliography

D. Applebaum and M. Riedle. Cylindrical Lévy processes in Banach spaces. *Proceedings of the London Mathematical Society*, 101(3):697–726, 2010.

S. Athreya and R. Tribe. Uniqueness for a class of one-dimensional stochastic pde's using moment duality. *Ann. Probab.*, 28(4):1711–1734, 2000.

S. Athreya, O. Butkovsky, and L. Mytnik. Strong existence and uniqueness for stable stochastic differential equations with distributional drift. *ArXiv e-prints*, January 2018.

D. Bakry. Semimartingales à deux indices. *Annales Scientifiques de l'Université de Clermont-Ferrand II. Mathématiques*, (20), 1982.

J.M. Ball. Strongly continuous semigroups, weak solutions, and the variation of constants formula. *Proceedings of the American Mathematical Society*, 63(2):370–373, 1977.

M. Barczy, Z. Li, and G. Pap. Yamada-Watanabe results for stochastic differential equations with jumps. *ArXiv e-prints*, November 2014.

M. T. Barlow. One dimensional stochastic differential equations with no strong solution. *Journal of the London Mathematical Society*, s2-26(2):335–347, 1982.

A. Barnosky. Distinguishing the effects of the red queen and court jester on miocene mammal evolution in the northern rocky mountains. *Journal of Vertebrate Paleontology*, 21(1):172–185, 2001.

N. H. Barton. Genetic hitchhiking. *Phil. Trans. Roy. Soc. London B*, 355:1553–1562, 2000.

N. H. Barton, A. M. Etheridge, and A. Sturm. Coalescence in a random background. *Ann. Appl. Probab.*, 14(2):754–785, 2004.

N. H. Barton, A. M. Etheridge, and A. Véber. A new model for evolution in a spatial continuum. *Electron. J. Probab.*, 15:162–216, 2010.

N. H. Barton, A. M. Etheridge, and A. Véber. Modelling evolution in a spatial continuum. *J. Stat. Mech.*, page PO1002, 2013.

J. Benton. The red queen and the court jester: Species diversity and the role of biotic and abiotic factors through time. *Science*, 323(5915):728–732, 2009.

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(11):e1004775, 2014.

J Bertoin and J-F Le Gall. Stochastic flows associated to a coalescent process. *Prob. Theor. Rel. Fields*, 126:261–288, 2003.

M Birkner. *Particle systems with locally dependent branching : long-time behaviour, genealogy and critical parameters*. PhD thesis, Goethe-Universität, Frankfurt am Main, 2003.

M. Birkner, J. Geiger, and G. Kersting. Branching processes in random environment—a view on critical and subcritical cases. In *Interacting stochastic systems*, pages 269–291. Springer, 2005.

J Blath, A M Etheridge, and M Meredith. Coexistence in locally regulated competing populations and survival of branching annihilating random walk. *Ann. Appl. Probab.*, 17:1474–1507, 2007.

J. Brossard. Deux notions équivalentes d'unicité en loi pour les équations différentielles stochastiques. In *Séminaire de Probabilités XXXVII*, pages 246–250. Springer, 2003.

A. S. Cherny. On the uniqueness in law and the pathwise uniqueness for stochastic differential equations. *Theory of Probability & Its Applications*, 46(3):406–419, 2002.

J. Chetwynd-Diggle and A. Etheridge. Superbrownian motion and the spatial lambda-fleming-viot process. *Electron. J. Probab.*, 23:36 pp., 2018.

E. Çınlar. *Probability and stochastics*, volume 261. Springer Science & Business Media, 2011.

D. Crisan. Superprocesses in a Brownian environment. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 460(2041):243–270, 2004. ISSN 1364-5021.

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(36):E5021–E5028, 2015.

G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2008. ISBN 9780521059800.

R. Dalang. Extending the martingale measure stochastic integral with applications to spatially homogeneous s.p.d.e.'s. *Electron. J. Probab.*, 4:29 pp., 1999. doi: 10.1214/EJP.v4-43.

R. Dalang and C. Mueller. Some non-linear s.p.d.e's that are second order in time. *Electron. J. Probab.*, 8:21 p., 2003. doi: 10.1214/EJP.v8-123.

R. Dalang and L. Quer-Sardanyons. Stochastic integrals for spde's: A comparison. *Expositiones Mathematicae*, 29(1):67 – 109, 2011. ISSN 0723-0869.

D A Dawson. Geostochastic calculus. *Can. J. Statistics*, 6:143–168, 1978.

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

C. R. Doering, C. Mueller, and P. Smereka. Interacting particles, the stochastic Fisher-Kolmogorov-Petrovsky-Piscounov equation, and duality. *Physica A Statistical Mechanics and its Applications*, 325:243–259, July 2003.

P. Donnelly and T. G. Kurtz. A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.*, 24:698–742, 1996.

P. Donnelly and T. G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.*, 27:166–205, 1999a.

P. Donnelly and T. G. Kurtz. Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.*, 9:1091–1148, 1999b.

R Durrett and W-T Fan. Genealogies in expanding populations. *Ann. Appl. Probab.*, 26:3456–3490, 2016.

H. J. Engelbert. On the theorem of T. Yamada and S. Watanabe. *Stochastics and Stochastic Reports*, 36(3-4):205–216, 1991.

A M Etheridge. *An introduction to superprocesses*, volume 20 of *University lecture notes*. Amer. Math. Soc., 2000.

A M Etheridge. Drift, draft and structure: some mathematical models of evolution. *Banach Center Publ.*, 80:121–144, 2008.

A. M. Etheridge and T. G. Kurtz. Genealogical constructions of population models. *Annals of Probability*, 2018.

A M Etheridge, A Véber, and F Yu. Rescaling limits of the spatial Lambda-Fleming-Viot process with selection. *arXiv:1406.5884*, 2014.

S. N. Ethier and T.G. Kurtz. *Markov processes: characterization and convergence*. Wiley, 1986.

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

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*, 1:143–174, 1947.

I. Gihman and A. V. Skorohod. Stochastic differential equations. In *The Theory of Stochastic Processes III*, pages 113–219. Springer, 1979.

J. Gillespie. Natural selection for variances in offspring numbers: a new evolutionary principle. *The American Naturalist*, 111(981):1010–1014, 1977.

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

J H Gillespie. Is the population size of a species relavant to its evolution? *Evolution*, 55:2161–2169, 2001.

J H Gillespie. *Population genetics: a concise guide*. The Johns Hopkins University Press, 2nd Edition, 2004.

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

A. Gupta. Stochastic model for cell polarity. *The Annals of Applied Probability*, 22(2): 827–859, 2012.

O Hallatschek and D Nelson. Gene surfing in expanding populations. *Theor. Pop. Biol.*, 73:158–170, 2008.

E. Huerta-Sanchez, R. Durrett, and C. D. Bustamante. Population genetics of polymorphism and divergence under fluctuating selection. *Genetics*, 178(1):325–337, 2008.

M. Hutzenthaler, Pfaffelhuber, and C. Printz. Stochastic averaging for multiscale markov processes with an application to branching random walk in random environment. *arXiv preprint arXiv:1504.01508*, 2018.

J. Jacod. Weak and strong solutions of stochastic differential equations. *Stochastics*, 3 (1-4):171–191, 1980.

I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

M Kimura. Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. *Genetics*, 39(3):280–295, 1954. ISSN 0016-6731.

W.H Kliemann, G. Koch, and F. Marchetti. On the unnormalized solution of the filtering problem with counting process observations. *IEEE Transactions on Information Theory*, 36(6):1415–1425, 1990.

S. M. Krone and C. Neuhauser. Ancestral processes with selection. *Theor. Pop. Biol.*, 51:210–237, 1997.

U. Kumar and M. Riedle. The stochastic Cauchy problem driven by a cylindrical Levy process. *ArXiv e-prints*, March 2018.

T. G. Kurtz. The Yamada-Watanabe-Engelbert Theorem for general stochastic equations and inequalities. *Electron. J. Probab.*, 12:951–965, 2007.

T. G. Kurtz. Equivalence of stochastic equations and martingale problems. In Dan Crisan, editor, *Stochastic Analysis 2010*, pages 113–130. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

T. G. Kurtz and P. Protter. Weak convergence of stochastic integrals and differential equations ii: Infinite dimensional case. In *Probabilistic models for nonlinear partial differential equations*, pages 197–285. Springer, 1996.

T. G. Kurtz and E. Rodrigues. Poisson representations of branching markov and measure-valued branching processes. *Ann. Probab.*, 39(3):939–984, 05 2011.

T.G. Kurtz. A limit theorem for perturbed operator semigroups with applications to random evolutions. *Journal of Functional Analysis*, 12(1):55–67, 1973.

T.G. Kurtz. Averaging for martingale problems and stochastic approximation. In *Applied Stochastic Analysis*, volume 177. Springer, Berlin, 1992.

T.G. Kurtz. Martingale problems for conditional distributions of Markov processes. *Elect. J. Prob.*, 3, paper 9, 1998.

T.G. Kurtz. Equivalence of stochastic equations and martingale problems. In *Stochastic analysis 2010*, pages 113–130. Springer, 2010.

T.G. Kurtz. Weak and strong solutions of general stochastic models. *Electron. Commun. Probab.*, 19:16 pp., 2014.

P.D. Lax. *Functional analysis*. Pure and applied mathematics. Wiley, 2002. ISBN 9780471556046.

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

V. B. Lidskiĭ. Non-selfadjoint operators with a trace. *Dokl. Akad. Nauk SSSR*, 125: 485–487, 1959. ISSN 0002-3264.

Wei Liu and Michael Röckner. *Stochastic partial differential equations: an introduction*. Springer, 2015.

Alessandra Lunardi. *Analytic semigroups and optimal regularity in parabolic problems*. Springer Science & Business Media, 2012.

Alessandra Lunardi. *Interpolation theory*, volume 16. Springer, 2018.

C. Ma. Construction of non-gaussian random fields with any given correlation structure. *Journal of Statistical Planning and Inference*, 139(3):780–787, 2009.

J Maynard Smith and J Haigh. The hitch-hiking effect of a favourable gene. *Genet. Res.*, 23:23–35, 1974.

R. Mikulevicius and B. Rozovskii. Normalized stochastic integrals in topological vector spaces. In *Séminaire de probabilités XXXII*, pages 137–165. Springer, 1998.

R Mikulevicius and B. Rozovskii. Martingale problems for stochastic pdes. *Stochastic partial differential equations: six perspectives*, 64:243–326, 1999.

L Miller. The effect of fluctuating selection in population genetics. Transfer of Status dissertation, 2012.

P A P Moran. Random processes in genetics. *Proc. Camb. Phil. Soc.*, 54:60–71, 1958.

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*, 102:519–546, 1995.

L Mytnik. *Superprocesses in random environments and related topics*. Ph D thesis, Technion, 1996a.

L Mytnik. Superprocesses in random environments. *Ann. Probab.*, 24:1953–1978, 1996b.

L. Mytnik, E. Perkins, and A. Sturm. On pathwise uniqueness for stochastic heat equations with non-lipschitz coefficients. *Ann. Probab.*, 34(5):1910–1959, 09 2006.

M. Nakashima. Branching random walks in random environment and super-brownian motion in random environment. *Ann. Inst. H. Poincaré Probab. Statist.*, 51(4): 1251–1289, 11 2015.

C. Neuhauser and S. M. Krone. Genealogies of samples in models with selection. *Genetics*, 145:519–534, 1997.

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

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

M. Ondreját. Uniqueness for stochastic evolution equations in banach spaces. *Dissertationes Mathematicae - Rozprawy Matematyczne*, 426:1–63, 01 2004.

C. Prévôt and M. Röckner. *A Concise Course on Stochastic Partial Differential Equations*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2007. ISBN 9783540707806.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

D Revuz and M Yor. *Continuous martingales and Brownian motion*. Springer, 1999.

M. Riedle. Stochastic integration with respect to cylindrical Lévy processes in Hilbert spaces: an L2 approach. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 17(01):1450008, 2014.

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

T. Rippl and A. Sturm. New results on pathwise uniqueness for the heat equation with colored noise. *Electron. J. Probab.*, 18:46 pp., 2013.

M. Röckner, B. Schmuland, and X. Zhang. Yamada-Watanabe Theorem for stochastic evolution equations in infinite dimensions. *Condensed Matter Physics*, 2008.

T Shiga. Stepping stone models in population genetics and population dynamics. In S Albeverio et al, editor, *Stochastic processes in physics and engineering.* D Reidel Publishing Company, 1988.

R. Stockbridge. Portfolio optimization in markets having stochastic rates. In Bozenna Pasik-Duncan, editor, *Stochastic Theory and Control*, pages 447–458, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

D. W. Stroock and S. R. S. Varadhan. On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Probability Theory*, pages 333–359, Berkeley, Calif., 1972. University of California Press.

A. Sturm. On convergence of population processes in random environments to the stochastic heat equation with colored noise. *Electron. J. Probab.*, 8:39 p., 2003.

S. Tappe. The Yamada-Watanabe Theorem for mild solutions to stochastic partial differential equations. *Electron. Commun. Probab.*, 18:13 pp., 2013.

J. E. Taylor. The genealogical consequences of fecundity variance polymorphism. *Genetics*, 182(3):813–837, 2009.

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

R. Tribe. Large time behavior of interface solutions to the heat equation with fisher-wright noise. *Prob. Th. Rel. Fields*, 102:289–311, 1995.

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

J B Walsh. An introduction to stochastic partial differential equations. In *École d'été de probabilités de Saint Flour*, volume 1180. Springer-Verlag, 1986.

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

T. Yamada and S. Watanabe. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11(1):155–167, 1971.

N. Young. *An Introduction to Hilbert Space.* Cambridge mathematical textbooks. Cambridge University Press, 1988. ISBN 9780521337175.

Appendix A

Markov Mapping Theorem

This Appendix contains several versions of the Markov Mapping Theorem used throughout the thesis. We aim to provide precise statements just as they appear in papers. First we recall some basic definitions and introduce the necessary notation. For a detailed account of this introductory material we refer to Lunardi (2012) and Ethier and Kurtz (1986), Chapter 1.

Let (E, d), (E_0, d_0) be a pair of complete, separable metric spaces (with metrics d and d_0, respectively). Let $B(E)$ be the space of bounded measurable functions on E. We notice that equipped with the usual supremum norm $\|\cdot\|_\infty$ it forms a Banach space. Let $C(E) \subset B(E)$ denote the subspace of continuous functions on E. A subspace A of $B(E) \times B(E)$ is a multivalued linear operator. It domain is given by $\mathcal{D} = \{f : (f, g) \in A\}$ and its range by $\mathcal{R} = \{g : (f, g) \in A\}$.

Definition A.1 (Dissipative operator)**.** *We say that the operator A is dissipative if for each $(f, g) \in A$ and $\lambda > 0$*

$$\|\lambda f - g\| \geq \lambda \|f\|.$$

Definition A.2 (Graph separable pre-generator)**.** *We say that an operator $A \subset B(E) \times B(E)$ is a pre-generator if it is dissipative and there exists a sequence of functions μ_n mapping E to the set of probability measures over $\mathcal{P}(E)$, and a sequence of $\lambda_n \in E$, such*

that for each $(f, g) \in A$

$$g(x) = \lim_{n\to\infty} \lambda_n \int_E (f(y) - f(x))\mu_n(x, \mathrm{d}y).$$

If in addition there exists a countable subset $\{f_k\} \subset \mathcal{D}(A) \bigcup C(E)$ *such that the graph of* A *is contained in the closure of the linear span of* (f_n, Af_n)*, we say that it is graph-separable.*

We notice that the generators of Markov process are graph-separable pre-generators. Let $D_E[0, \infty)$ denote the space if càdlàg functions and $M_E[0, \infty)$ denote the space of Borel measurable functions from $[0, \infty)$ taking values in E.

We are ready to state two versions of the Markov Mapping Theorem.

Theorem A.3 (Kurtz (2011), Theorem 1)**.** *Suppose that* $B \subset C(E) \times C(E)$ *is a graph separable pre-generator. Moreover, assume that* $\mathcal{D}(B)$ *is closed under multiplication and separates points in* E*. Let* (E_0, r_0) *be a complete, separable metric space,* $\gamma : E \to E_0$ *be Borel measurable, and let* α *be a transition function from* E_0 *into* E *(that is,* $E_0 \ni y \mapsto \alpha(y, \cdot) \in \mathcal{P}(E)$ *is Borel measurable) satisfying* $\alpha(y, \gamma^{-1}(y)) = 1$*. Define*

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

Let $\mu_0 \in \mathcal{P}(E_0)$*, and define* $\nu_0 = \int_{\mathbb{R}} \alpha(y, \cdot)\mu_0(\mathrm{d}y)$*. If* $\widetilde{Y}$ *is a solution of the martingale problem for* (C, μ_0)*, then there exists a solution* X *of the martingale problem for* (B, ν_0) *such that* $\widetilde{Y}$ *has the same distribution on* $\mathcal{M}_{E_0}[0, \infty)$ *as* $Y = \gamma \circ X$*. If* $\tilde{Y}$ *(and hence* Y*) has a modification with sample paths in* $\mathcal{D}_E([0, \infty))$*, then the modified* $\tilde{Y}$ *and* Y *have the same distribution on* $\mathcal{D}_E([0, \infty))$*.*

The following extension plays a crucial role in Chapter 4.

Theorem A.4 (Kurtz and Rodrigues (2011), Theorem A.15)**.** *Let* $A \subset \overline{C}(E) \times C(E)$ *and let* ψ *be a continuous function taking values in* $\mathbb{R}$ *such that* $\psi \geq 1$*. Suppose that for each* $f \in \mathcal{D}(A)$ *there exists a constant* $c_f > 0$ *such that*

$$|Af(x)| \leq c_f \psi(x).$$

Let A_0 be defined as

$$A_0 f(x) = \frac{Af(x)}{\psi(x)}.$$

Suppose that A_0 is graph-separable pre-generator and suppose that $\mathcal{D}(A) = \mathcal{D}(A_0)$ is closed under multiplication and separating. Let $\gamma : E \to E_0$ be Borel measurable, and let α be a transition function from E_0 into E satisfying $\alpha(y, \gamma^{-1}(y)) = 1$. Assume that for each $y \in E_0$

$$\widetilde{\psi} = \int_E \psi(y, z)\alpha(y, \mathrm{d}z) < \infty.$$

and define

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

Let $\mu_0 \in \mathcal{P}(E_0)$, and define $\nu_0(y) = \int_{\mathbb{R}} \alpha(y, \cdot)\mu_0(\mathrm{d}y)$.

1. *Let $\tilde{Y}$ be a solution of a martingale problem for (C, μ_0). Assume that it satisfies the moment condition*

$$\int_0^t \mathbb{E}\left[\widetilde{\psi}(\widetilde{Y}(s))\mathrm{d}s\right] < \infty \quad \forall t \geq 0.$$

   Then there exists a solution X of the martingale problem for (A, ν_0) such that $\tilde{Y}$ has the same distribution on $\mathcal{M}_{E_0}[0, \infty)$ as $\widetilde{Y} = \gamma \circ Y$.

2. *If, in addition, uniqueness holds for the martingale problem for (A, ν_0), then uniqueness holds for the martingale problem for (C, μ_0).*

Appendix B

Numerical Simulations

This Appendix is devoted to a more precise description of numerical results of Section 1.4. Section B.1 is devoted to a description of our simulations. Section B.2 contains technical details of the implementation.

B.1 Detailed description of simulations

B.1.1 A general description

In this subsection we provide a more detailed description of the simulation procedure.

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 1.4 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 allows 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 initialized. 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 initialized 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 event: two types of reproduction event (neutral and selective), migration events and environmental events.

Whenever a reproduction event takes place in deme α, two individuals are sampled uniformly from deme α. One of the chosen 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 picked 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 chosen according to the uniform distribution. A single individual is chosen uniformly from each of the 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 1.4, 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.

B.2 Implementation

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