The Annals of Applied Probability
2004, Vol. 14, No. 1, 188–214

SURVIVAL AND EXTINCTION IN A LOCALLY REGULATED POPULATION

BY A. M. ETHERIDGE[1]

University of Oxford

Bolker and Pacala recently introduced a model of an evolving population in which an individual's fecundity is reduced in proportion to the "local population density." We consider two versions of this model and prove complementary extinction/persistence results, one for each version. Roughly, if individuals in the population disperse sufficiently quickly relative to the range of the interaction induced by the density dependent regulation, then the population has positive chance of survival, whereas, if they do not, then the population will die out.

1. Introduction. For over a century, branching processes have been used to model the evolution of biological populations. The classical *Galton–Watson process* models the total population size, and if, in addition, during their lifetime individuals are assumed to follow independent Brownian motions (or random walks) then we arrive at *branching Brownian motion* (or branching random walk). This process can then be used to model the evolution of a population that is dispersed in space. For large populations, Feller (1951) observed that, in suitable units, one can approximate the evolution of the population size by a one-dimensional diffusion. This idea can be extended to the spatial setting, leading to the *Dawson–Watanabe superprocess* (also known as superBrownian motion).

However, all such models have long been known to be deficient. The most obvious problem is that for finite populations, branching processes do not predict a stable population size: either the population will die out or it will grow without bound. In the spatial setting, one attempt to combat this is to consider infinite populations, evolving in the whole of d-dimensional Euclidean space. Although in at least three spatial dimensions, the branching Brownian motion and superBrownian motion models both have a nontrivial equilibrium, most biological populations live in *two* spatial dimensions and in (one and) two dimensions, the branching Brownian motion and superprocess models *do not* have an equilibrium distribution. Worse, if not extinct, at large times, the process will form "clumps" of arbitrarily large density and extent [Felsenstein (1975)].

A drastic solution to the first of these shortcomings is to condition the total population to be constant (or otherwise exogenously specified). This leads to the *Wright–Fisher* and *Moran models* or, in the superprocess setting, to the

---

Received September 2001; revised March 2003.

[1] Supported in part by EPSRC GR/L67899 and by an EPSRC Advanced Fellowship.

AMS 2000 subject classifications. 60J80, 60J85, 60J70, 60K35.

Key words and phrases. Interacting superprocess, regulated population, extinction, persistence.

Fleming–Viot process. Unfortunately, for large populations, this does not combat the problem of clumping and, moreover, we still require a satisfactory model for the population size. An alternative and popular, approach is to restrict the population to live in discrete "demes," each of which is of a fixed size. This is the so-called *stepping-stone model* [Kimura (1953)]. However, for populations evolving in continua, this restriction to discrete demes is unnatural and the fixed size of the demes may disguise important effects arising from random fluctuations in the local population density [see Barton, Depaulis and Etheridge (2002) for a discussion of this issue]. In particular, although the stepping stone model may give good predictions, the parameters of the model should be replaced by "effective parameters" whose values are determined by detailed local structure.

Rather than having an exogenously specified (global) population size, one expects that a real population should be regulated by local rules. It is natural, for example, to suppose that an individual living in a crowded region will have fewer offspring reach maturity than one living in a sparsely populated region. In one spatial dimension, Mueller and Tribe (1994) study a stochastic partial differential equation that can be used to model the *population density*, defined as a positive function on $\mathbb{R}$, for such a population. In this model, the reproductive success of an individual decreases linearly with the local population density. More precisely, they study the following equation:

$$u_t = \frac{1}{6}u_{xx} + u(\theta - u) + u^{1/2}\dot{W}, \qquad t > 0, x \in \mathbb{R}, \theta > 0, \tag{1}$$

where $\dot{W}$ is space–time white noise and $u(0, x)$ is a positive function, of compact support. (The constant $1/6$ in front of the dispersal term arises from their construction of the solution as a limit of rescaled contact processes and is of no greater significance.) They say that the process *survives* if, for all $t > 0$, $u(t, 0) \not\equiv 0$. Mueller and Tribe proved the following result.

THEOREM 1.1 [Mueller and Tribe (1994)]. *There exists a constant* $\theta_c > 0$, *not depending on* $u(0, \cdot)$, *such that*:

(i) *if* $\theta < \theta_c$, *then* $\mathbb{P}[u(t, x) \text{ survives}] = 0$,
(ii) *if* $\theta > \theta_c$, *then* $\mathbb{P}[u(t, x) \text{ survives}] > 0$.

This model cannot be extended to higher spatial dimensions, for then, if a solution to equation (1) were to exist, it would be distribution valued [Walsh (1986)] which renders the nonlinear term meaningless. Moreover, for the analogous rescaling of the contact process in dimensions bigger than one, the limit process degenerates to superBrownian motion [Durrett and Perkins (1999)].

In Bolker and Pacala (1997) a model is proposed, based upon branching random walk, in which the mean number of offspring of an individual at x is again governed by the "local population density," but this is now defined as a weighted sum of the entire population, with weights depending only on distance from x.

For example, an individual's fecundity could be reduced according to the size of the population in an open neighborhood around her. The model makes sense in any spatial dimension and is easily adapted to the branching Brownian motion, stepping stone and superprocess settings.

Our results are concerned with extinction and persistence of models of the Bolker–Pacala type. We shall prove an extinction result for a superprocess version of the model and a survival result for a stepping stone version. Although we have shamelessly chosen the simplest cases, we believe that the key ideas for proving the corresponding results for all versions of the Bolker–Pacala model are contained in these examples. Our decision to present an incomplete result followed the realization that not only was the paper becoming repetitive, but also the notation and technicalities were spiralling out of control and obscuring the essentially simple structure of the proofs.

Before stating our result, we must be more precise about the model. First we describe *the superprocess version of the Bolker–Pacala model*. From now on, angle brackets are used to denote integration, thus for a function ϕ and a measure μ,

$$\langle \phi, \mu \rangle \equiv \int_{\mathbb{R}^d} \phi(x)\mu(dx).$$

The state of the population at time t is described by a measure, X_t. Writing $\mathbb{P}$ for the distribution of the process $\{X_t\}_{t\geq 0}$, for suitable test functions ϕ,

$$\begin{aligned} M_t(\phi) &\triangleq \langle \phi, X_t\rangle - \langle \phi, X_0\rangle - \int_0^t \left\langle \tfrac{1}{2}\Delta\phi, X_s\right\rangle ds \\ &\quad - \int_0^t \left\langle \alpha\big(M - \langle h(\|x-y\|), X_s(dy)\rangle\big)\phi(x), X_s(dx)\right\rangle ds \end{aligned} \tag{2}$$

is a $\mathbb{P}$-martingale with quadratic variation

$$[M(\phi)]_t = \int_0^t \gamma \langle \phi^2, X_s\rangle\, ds.$$

This differs from the Dawson–Watanabe superprocess martingale problem only in the *local regulation* term involving α, M and h. Here α, γ and M are positive constants that determine, respectively, the rate of reproduction in the population (in the branching random walk model this corresponds to the inverse of the mean inter-generation time), the variance of the number of offspring of individuals in the population and the *carrying capacity* of the *neighborhood* of an individual. The "neighborhood" is determined by the function $h:\mathbb{R}_+ \to \mathbb{R}_+$, that we shall refer to as the *interaction kernel*. If h is an indicator function, $\chi_{[0,R]}$ say, then the neighborhood of an individual is simply the ball of radius R centred on that individual. More generally, one might expect that the impact on the reproductive success of an individual at x of an individual at y should decrease with $\|x-y\|$ (the physical separation of the individuals), and so we shall take h to be a monotone decreasing function on $\mathbb{R}_+$. (Although we shall take h to be bounded, one can also allow singular interaction kernels, see Remark 3.4.) One cannot expect to construct the process (nontrivially) for arbitrary initial measures, there has to be some control over the rate at which they "grow" at infinity. We shall give a formal definition of the process in Section 2 (Definitions 2.1–2.3) where, in particular, an appropriate state space for the process, *tempered measures*, will be described in detail. This class includes Lebesgue measure. In the case of infinite initial measures, we impose two additional assumptions. First we require a decay condition on h to prevent instant catastrophe for the population [equation (3) below]. Second, since we have not proved uniqueness of the solution to the martingale problem for infinite initial measures, we restrict ourselves to those solutions that are obtained as rescalings of branching particle systems as described in Definition 2.3.

NOTATION 1.2. An *interaction kernel* is a bounded, monotone decreasing function $h : \mathbb{R}_+ \to \mathbb{R}_+$. In the context of the superprocess version of the Bolker–Pacala model with an infinite initial measure, h is assumed to satisfy the additional condition

$$\int_0^\infty h(r) r^{d-1}\, dr < \infty. \tag{3}$$

Now suppose that we are working with a subdivided population, restricted to live on the lattice $\mathbb{Z}^d$. We denote the size of the population in deme $i \in \mathbb{Z}^d$ at time t by $X_t(i)$. As above, α and M will be positive constants, but now the Laplacian will be replaced by a migration matrix $(m_{ij})_{i,j\in\mathbb{Z}^d}$, and the interaction kernel, h, is replaced by an *interaction matrix*, $(\lambda_{ij})_{i,j\in\mathbb{Z}^d}$. Both the migration and interaction matrices have only nonnegative entries and are assumed to be nondiagonal. We shall also take m_{ij} and λ_{ij} to be functions of $j - i$ alone and assume that

$$\sum_j m_{ij}, \sum_j \lambda_{ij} < \infty. \tag{4}$$

Notice that this second condition reflects equation (3).

DEFINITION 1.3. We shall say that the population follows *the stepping stone version of the Bolker–Pacala model* if its evolution is described by the following system of stochastic differential equations:

$$\begin{aligned} dX_t(i) &= \sum_{j\in\mathbb{Z}^d} m_{ij}\big(X_t(j) - X_t(i)\big)\, dt \\ &\quad + \alpha\Bigg(M - \sum_{j\in\mathbb{Z}^d} \lambda_{ij} X_t(j)\Bigg) X_t(i)\, dt + \sqrt{\gamma X_t(i)}\, dB_t^{(i)}, \end{aligned} \tag{5}$$

where $\{\{B_t^{(i)}\}_{t\geq 0}\}_{i\in\mathbb{Z}^d}$ is a collection of independent Brownian motions.

Much of the elegant mathematics that arises in the study of branching process models stems from the *branching property*: the distribution of the process started from $\mu_1 + \mu_2$ is the same as that of the sum of two independent copies of the process started from μ_1 and μ_2, respectively. The interaction (local regulation) in (2) and (5) destroys this convenient property as it is no longer the case that, once born, individuals in the population evolve independently of one another. As a result, the distribution can no longer be characterized in terms of a partial differential equation. The form of the nonlinearity destroys monotone dependence on the initial condition (at least for nonconstant h) and ensures that the moment equations are not closed (the equation for the first moment involves the second and so on). Consequently, analytic results are hard to obtain and our knowledge of the process is heavily dependent upon computer simulations.

In addition to simulations, Bolker and Pacala (1997, 1999) use moment closure to study the process. The idea is that one should ignore the higher order moments (i.e., set them equal to zero) and thus "close off" the moment equations. In fact, they set the third moment to be zero. In their model, after forty generations or so, they remark that their simulations appear to have arrived at a "statistical equilibrium" for the process and they use the moment equations to discuss this supposed stationary distribution. There are of course reasons to be wary. The process is a spatial analogue of the following *Feller diffusion with logistic growth*,

$$dx_t = \alpha x_t(M - \lambda x_t)\,dt + \sqrt{\gamma x_t}\,dB_t, \tag{6}$$

where B_t is Brownian motion and λ is yet another positive constant. This diffusion will, with probability one, hit zero in finite time. (This is, of course, in stark contrast to the corresponding *deterministic* equation.) For equation (6), this moment closure does not predict a stationary distribution, but neither does it correctly predict the long time behavior of the process.

DEFINITION 1.4. 1. We shall say that the superprocess (resp. stepping stone) version of the model *suffers local extinction* if for each $\varepsilon > 0$, $\mathbb{P}[\langle \chi_{B(0,1)}, X_t \rangle > \varepsilon] \to 0$ (resp. $\mathbb{P}[X_t(0) > \varepsilon] \to 0$) as $t \to \infty$, where $B(0, 1)$ is the ball of radius one centered on the origin.

2. We shall say that the superprocess (resp. stepping stone) version of the model *survives* if there exist $\varepsilon, \delta > 0$ such that $\liminf_{t\to\infty} \mathbb{P}[\langle \chi_{B(0,1)}, X_t \rangle > \varepsilon] > \delta$ (resp. $\liminf_{t\to\infty} \mathbb{P}[X_t(0) > \varepsilon] > \delta$).

If the population is evolving in a compact space, then we expect that it should die out in finite time, but for infinite environments, by analogy with equation (1), one might hope that for some choices of parameter the process will survive. In this setting, the form of the interaction kernel will play a critical rôle. If the interaction is too "long range" (measured from the perspective of the rate of dispersal of individuals in the population), then the population will die out. This is true even for *infinite* initial measures. Roughly, for the population to survive, families must disperse fast enough (in relation to the interaction kernel) that they colonize new sites before being killed by overcrowding. Law, Murrell and Dieckmann (2003) independently arrived at this conclusion using simulations of the population. Although our primary interest is in two dimensions, our results remain valid for *all* dimensions. More precisely, we prove the following theorem.

THEOREM 1.5. 1. *For each fixed interaction kernel* h *and* $\gamma, K > 0$ *there exists* $\alpha_0 = \alpha_0(K, \gamma, h)$ *such that for* $\alpha > \alpha_0$, *the superprocess version of the Bolker–Pacala model with parameters* $(h, K/\alpha, \alpha, \gamma)$ *started from any finite initial measure dies out in finite time. If* h *also satisfies* (3), *then when started from any tempered initial measure* (*with* $p > d$) *the process with these parameters suffers local extinction.*

2. *Let* $\alpha > 0$ *be fixed.*

(a) *If the interaction kernel* h *is such that* $r^{2-\delta}h(r)$ *is unbounded for some* $\delta > 0$, *then for each fixed* $\gamma > 0$, *there is an* $M_0 > 0$ *such that for* $M < M_0$ *the superprocess version of the Bolker–Pacala model with parameters* (h, M, α, γ) *started from any finite initial measure dies out in finite time. If* h *also satisfies* (3), *so that in particular* $d = 1$, *then when started from any tempered initial measure* (*with* $p > 1$) *the process with these parameters suffers local extinction.*

(b) *Suppose that the population* $\{\{X_t(i)\}_{i\in\mathbb{Z}^d}\}_{t\geq 0}$ *evolves according to the stepping stone version of the Bolker–Pacala model, then if* $m_{ij} > c\lambda_{ij}$, *for some* $c > 0$, *then there exists* $M_1 > 0$ *such that for* $M > M_1$ *the process survives for all time with* (*strictly*) *positive probability* (*started from any nontrivial initial condition*).

REMARK 1.6. 1. For *infinite* initial measures, because we have not proved uniqueness of the solution to the martingale problem, we define the superprocess version of the model to be the limit of rescaled branching particle systems (see Definition 2.3).

2. The results in (2) in some sense complement one another. Suppose that we approximate our continuous (superprocess) version of the model by a stepping stone model in which demes are separated by distance r. The coefficients in the migration matrix represent the rates at which individuals migrate between demes and will scale therefore (in the case of Brownian spatial motion) with r^{-2}. Roughly then, the condition $m_{ij} > c\lambda_{ij}$ is equivalent to $r^{-2} > ch(r)$. In other words $r^2h(r)$ is bounded. [In two spatial dimensions, the integrability condition (3) ensures this.] If the spatial motion of individuals were a *stable process* of index $\beta \in (0, 2)$ rather than Brownian motion, then this condition would be replaced by the condition that $r^{\beta}h(r)$ be bounded.

3. The integrability condition (3) is included specifically to allow us to construct the process from a translation invariant initial state such as the Lebesgue

measure. If the process is to survive then (at least in two dimensions) one might hope for a translation invariant stationary distribution (see Section 3). If $r^{2-\delta}h(r)$ is unbounded, then there is a problem even constructing the process started from such a distribution which suggests that we should be able to set $M_0 = \infty$ in 2(a). Our methods do not seem to be sufficiently powerful to capture this.

To prove extinction, we shall dominate the population by a process of "clusters," with the number of clusters evolving according to a subcritical branching process. This is rather similar to the proof of the corresponding part of Theorem 1.1, although the fact that our process involves interactions between individuals at different spatial locations means that we cannot crib directly from Mueller and Tribe (1994). It is also this "nonlocal" nature of the interaction that invalidates the comparison results that are combined with a comparison with an *oriented percolation* process in Mueller and Tribe (1994) to prove the second half of Theorem 1.1, necessitating a somewhat different approach to the comparison here. (It is because the case of diagonal interaction matrices is essentially covered by the Mueller and Tribe result that we exclude it here.) Nevertheless, to prove the second half of Theorem 1.5 we *do* compare the process to an oriented percolation process, that is, to a discrete time version of the *contact process*. The reader is refered to Durrett (1995) for very general comparison techniques for interacting particle systems.

Extinction in the superprocess setting is associated with *clumping* of the process. The density dependent regulation term in the Bolker–Pacala model is introduced to overcome the clumping and we shall see in Section 3, via an heuristic argument, that survival of the process is a reflection of successful eradication of clumping.

The rest of this article is laid out as follows. In Section 2 we introduce notation and give a precise description of the superprocess version of the model. In Section 3 we present an heuristic argument to show why we should expect Theorem 1.5 to hold. In passing we also record some facts about the Dawson–Watanabe superprocess that dictate the restriction on the possible forms of singular interaction kernels alluded to above. We recall these facts without proof. The reader seeking an introduction to superprocesses (with and without interactions) is referred to Dawson (1993), Etheridge (2000), Le Gall (1999) and Perkins (2000). In Section 4, we illustrate the method of proof of extinction by applying it to the diffusion (6). This will also be useful in explaining why our extinction proof breaks down if the parameter M is too large. In Section 5 we prove the extinction result via a comparison with a subcritical (age-dependent) branching process. A sketch of the proof of survival for subdivided populations is provided in Section 6. Finally, Section 7 is a brief discussion of the result and some avenues for future research.

2. Notation and description of the process.

In this section we fix our notation and give a precise description of the superprocess version of the Bolker–Pacala model.

First some notation. We write $\mathcal{M}_F(\mathbb{R}^d)$ for the space of *finite measures* on $\mathbb{R}^d$, assumed to be endowed with the weak topology. The space of *p-tempered measures* on $\mathbb{R}^d$, that is, the space of locally finite measures, μ, such that

$$\int_{\mathbb{R}^d} \phi_p(x)\mu(dx) \triangleq \int_{\mathbb{R}^d} \frac{1}{(1+\|x\|^2)^{p/2}}\mu(dx) < \infty, \tag{7}$$

will be denoted $\mathcal{M}_p(\mathbb{R}^d)$. It is endowed with a natural coarsening of the weak topology that we shall call the $C_p(\mathbb{R}^d)$-topology. $C_p(\mathbb{R}^d)$ is our notation for the space of functions of the form $f + \beta\phi_p$, where f is a continuous function with compact support on $\mathbb{R}^d$, ϕ_p is the function implicitly defined in (7) and $\beta \in \mathbb{R}$. A sequence of measures $\{\mu_n\}_{n\geq 1}$ in $\mathcal{M}_p(\mathbb{R}^d)$ converges in the $C_p(\mathbb{R}^d)$-topology to the limit μ if

$$\langle \phi, \mu_n\rangle \overset{n\to\infty}{\longrightarrow} \langle \phi, \mu\rangle \qquad \forall\, \phi \in C_p(\mathbb{R}^d).$$

The space of càdlàg (right continuous with left limits) mappings from $\mathbb{R}_+$ to $\mathcal{M}_p(\mathbb{R}^d)$ will be denoted by $D(\mathbb{R}_+, \mathcal{M}_p(\mathbb{R}^d))$, or more concisely by Ω, as (for $p > d$) this will provide a state space for the process started from infinite initial measures. We abuse notation and use the same notation for the space $D([0,\infty), \mathcal{M}_F(\mathbb{R}^d))$, which is the state space for the process started from a finite initial measure. We always assume that Ω is endowed with the natural filtration, $\{\mathcal{F}_t\}_{t\geq 0}$.

We shall denote our process by $\{X_t\}_{t\geq 0}$, and characterize its distribution as the solution to a martingale problem.

First suppose that we are interested in finite populations. The model is specified in terms of a bounded positive function $h : \mathbb{R}_+ \to \mathbb{R}_+$ and three positive constants, M, α and γ. We use Δ to denote the Laplacian on $\mathbb{R}^d$ and $\mathcal{D}(\Delta)$ for its domain.

DEFINITION 2.1. For $\mu \in \mathcal{M}_F(\mathbb{R}^d)$, the measure $\mathbb{P}_\mu$ solves the *BP-martingale problem* with parameters (h, M, α, γ) if $\mathbb{P}_\mu[X_0 = \mu] = 1$ and for positive, bounded functions $\phi \in \mathcal{D}(\Delta)$ of compact support,

$$\begin{aligned} M_t(\phi) &\triangleq \langle \phi, X_t\rangle - \langle \phi, X_0\rangle - \int_0^t \big\langle \tfrac{1}{2}\Delta\phi, X_s\big\rangle\, ds \\ &\quad - \int_0^t \big\langle \alpha\big(M - \langle h(\|x-y\|), X_s(dy)\rangle\big)\phi(x), X_s(dx)\big\rangle\, ds \end{aligned} \tag{8}$$

is a $\mathbb{P}_\mu$ ($\mathcal{F}_\cdot$-) martingale with quadratic variation

$$[M(\phi)]_t = \int_0^t \gamma \langle \phi^2, X_s\rangle\, ds. \tag{9}$$

In this case of a finite population, a process with the required distribution can be obtained from the superBrownian motion using Dawson's Girsanov transform, and, indeed, precisely this model is used as an example of the use of the transform in Dawson (1993). Uniqueness of the solution to the martingale problem follows from the version of Dawson's Girsanov transform provided in Evans and Perkins (1994). This in turn guarantees that, when started from a finite initial measure, the process has the strong Markov property, a fact that we shall need in our proof of extinction.

In order to construct the process from an infinite initial measure, we require a *decay condition* on h, that ensures that the population does not suffer an instant catastrophe.

DEFINITION 2.2. Suppose that the function h satisfies the additional condition

$$\int_0^\infty h(r) r^{d-1}\, dr < \infty. \tag{10}$$

For $\mu \in \mathcal{M}_p(\mathbb{R}^d)$ and $p > d$, the measure $\mathbb{P}_\mu$ solves the *BP-martingale problem* with parameters (h, M, α, γ) if $\mathbb{P}_\mu[X_0 = \mu] = 1$ and for positive, functions $\phi \in \mathcal{D}(\Delta) \cap C_p(\mathbb{R}^d)$, the expression (8) is a $\mathbb{P}_\mu$ ($\mathcal{F}_\cdot$-) martingale with quadratic variation given by (9).

For infinite populations the Girsanov transform is no longer applicable, but one can construct the process as a weak limit of rescaled (interacting) branching particle systems by obvious analogy with the corresponding construction for the classical superprocess. Our proof of the first part of Theorem 1.5 will exploit the approximating branching particle system. For technical reasons, we always assume that the initial measures for the approximating systems are finite, even if their limit is infinite. Since we have not proved uniqueness of the solution to the martingale problem in the case of an infinite initial measure, we must introduce the caveat that our result only applies to solutions constructed in this way (see, however, Section 7). We therefore make the following definition.

DEFINITION 2.3. We shall say that the population $\{X_t\}_{t\geq 0}$ evolves according to the superprocess version of the Bolker–Pacala model with parameters (h, M, α, γ) if either:

1. $X_0 \in \mathcal{M}_F(\mathbb{R}^d)$ (with probability one) and the distribution of $\{X_t\}_{t\geq 0}$ is the unique solution of the BP-martingale problem with parameters (h, M, α, γ), or
2. $X_0 \in \mathcal{M}_p(\mathbb{R}^d)$ for some $p > d$ and the distribution of $\{X_t\}_{t\geq 0}$ is the solution to the BP-martingale problem obtained by the limiting procedure described below.

We now, very briefly, describe the construction of the process as a limit of branching particle systems. We pave the way for Sections 4 and 5 by constructing a *pair* of processes $\{X_t, Y_t\}_{t\geq 0}$, in such a way that $\{Y_t\}_{t\geq 0}$ is a supercritical superprocess and $\{X_t\}_{t\geq 0}$ is a "subpopulation" of $\{Y_t\}_{t\geq 0}$ whose distribution solves the BP-martingale problem.

The process $\{Y_t\}_{t\geq 0}$ is obtained as the limit of a sequence of purely atomic measure valued processes $\{Y_t^n\}_{t\geq 0}$. We think of Y_t^n as a population in which each individual is represented by an atom of mass $1/n$. The evolution of $\{Y_t^n\}_{t\geq 0}$ is as follows. Each individual has an independent exponentially distributed lifetime, with parameter $\gamma n/\sigma^2$, during which she moves around in space according to a Brownian motion and at the end of which she leaves behind, at the location where she dies, a random number of offspring with mean $1 + \alpha M/n + O(1/n^2)$ and variance $\sigma^2 + O(1/n)$. We augment this process by assigning an independent uniform [0, 1] random variable to each individual of $\{Y_t^n\}_{t\geq 0}$.

The process $\{X_t^n\}_{t\geq 0}$ is a subpopulation of $\{Y_t^n\}_{t\geq 0}$. To describe the process, let us assign the label 1 or 0 to each individual in the $Y_\cdot^n$-population, according to whether or not she is a member of the $X_\cdot^n$-population. At time zero all individuals have label 1. If an individual has label 0, then so do all her descendants. If an individual with label 1 dies at the spatial position x at time τ, then each of her descendants has label 0 or 1 according as their uniform random variable, U, satisfies $U < \frac{\alpha}{n}\langle h(x-y), X_{\tau-}^n(dy)\rangle$ or $U > \frac{\alpha}{n}\langle h(x-y), X_{\tau-}^n(dy)\rangle$.

Notice that the net effect of this is that if an individual of the X-population dies at time τ and spatial position x, her mean number of descendants (in the X-population) is

$$1 + \frac{\alpha}{n}\big(M - \langle h(x-y), X_{\tau-}^n(dy)\rangle\big) + O\left(\frac{1}{n^2}\right)$$

and the variance of the number of offspring is $\sigma^2 + O(1/n)$.

Provided that

$$X_0^n \overset{w}{\to} X_0 \qquad \text{as } n \to \infty,$$

the càdlàg measure-valued processes $\{X_t^n\}_{t\geq 0}$ converge (weakly) to a solution to the BP-martingale problem as $n \to \infty$. Notice that the processes $\{Y_t^n\}_{t\geq 0}$ converge to a supercritical superprocess. This construction therefore guarantees that, provided $X_0 \in \mathcal{M}_p(\mathbb{R}^d)$, we will have $X_t \in \mathcal{M}_p(\mathbb{R}^d)$ for all $t > 0$.

We omit the details of the construction as they are completely standard, but let us remark that one has to work directly with particle systems that evolve in $\mathcal{M}_p(\mathbb{R}^d)$, rather than follow the usual route of partitioning the infinite initial measure into a countable number of finite pieces and summing independent copies of the superprocess started from these finite measures. This is once again because the branching property is destroyed by our interaction. Details of particle system based constructions of superprocesses can be found in our references on superprocesses. General results on convergence of stochastic processes can be found in Ethier and Kurtz (1986).

3. Rescaling and heuristics. In this section we rescale the model according to a rescaling under which the classical (critical) Dawson–Watanabe superprocess is invariant (up to a change in initial condition). We discuss the (heuristic) implications of the rescaling at the end of the section. Not only will it explain the critical rôle of the asymptotic behavior of $\theta^2 h(\theta)$ in determining extinction/survival of the population, but we will also see that, in two spatial dimensions, survival of the process reflects eradication of clumping.

To this end, we define a new process X^θ by

$$\langle \phi, X_t^\theta \rangle \triangleq \left\langle \frac{1}{\theta^2}\phi\left(\frac{x}{\theta}\right), X_{\theta^2 t}(dx)\right\rangle, \tag{11}$$

for all $\phi \in \mathcal{D}(\Delta)$ for which the integral is defined.

REMARK 3.1. In order for our initial condition to be invariant under this rescaling one typically chooses X_0 to be Lebesgue measure in two dimensions or, more generally, to have density $|x|^{2-d}$ with respect to Lebesgue measure in dimension d. Notice that for this choice of X_0, if $h(r) \geq 1/r^2$ as $r \to \infty$ then $\langle h(x, y), X_0(dy)\rangle$ is infinite and there are problems even constructing the superprocess version of our model.

LEMMA 3.2. *The distribution of the process* $\{X_t^\theta\}_{t\geq 0}$ *solves the BP-martingale problem* (*with initial condition* X_0^θ *and*) *with parameters* $(h^\theta, M, \theta^2\alpha, \gamma)$, *where*

$$h^\theta(r) = \theta^2 h(\theta r).$$

PROOF. It is convenient to write

$$\phi^\theta(x) = \frac{1}{\theta^2}\phi\left(\frac{x}{\theta}\right).$$

The proof is an elementary manipulation of the martingale characterization of the distribution of X_t: substituting into (8) we see that

$$\begin{aligned} &\langle \phi, X_t^\theta\rangle - \langle \phi, X_0^\theta\rangle - \int_0^{\theta^2 t} \langle \tfrac{1}{2}\Delta\phi^\theta, X_s\rangle\, ds \\ &\qquad - \int_0^{\theta^2 t} \langle \alpha(M - \langle h(\|x-y\|), X_s(dy)\rangle)\phi^\theta(x), X_s(dx)\rangle\, ds \end{aligned} \tag{12}$$

is a $\mathbb{P}_\mu$ ($\mathcal{F}_\cdot$-) martingale with quadratic variation

$$\int_0^{\theta^2 t} \gamma\langle (\phi^\theta)^2, X_s\rangle\, ds.$$

First observe that

$$\begin{aligned}
&\int_0^{\theta^2 t} \langle \alpha(M - \langle h(\|x-y\|), X_s(dy)\rangle)\phi^\theta(x), X_s(dx)\rangle\, ds \\
&\quad = \int_0^t \left\langle \alpha\theta^2 \left(M - \left\langle \frac{1}{\theta^2} h^\theta\left(\left\|\frac{x}{\theta} - \frac{y}{\theta}\right\|\right), X_{\theta^2 s}(dy)\right\rangle\right) \frac{1}{\theta^2}\phi\left(\frac{x}{\theta}\right), X_{\theta^2 s}(dx)\right\rangle ds \\
&\quad = \int_0^t \langle \alpha\theta^2 (M - \langle h^\theta(\|x-y\|), X_s^\theta(dy)\rangle)\phi, X_s^\theta(dx)\rangle\, ds.
\end{aligned}$$

Now change variable in the other integrals in (12) and use that

$$\Delta\phi^\theta(x) = \frac{1}{\theta^4}\Delta\phi\left(\frac{x}{\theta}\right),$$

to see that

$$\begin{aligned}
&\langle \phi, X_t^\theta\rangle - \langle \phi, X_0^\theta\rangle - \int_0^t \langle \tfrac{1}{2}\Delta\phi, X_s^\theta\rangle\, ds \\
&\qquad - \int_0^t \langle \theta^2\alpha(M - \langle h^\theta(\|x-y\|), X_s^\theta(dy)\rangle)\phi(x), X_s^\theta(dx)\rangle\, ds
\end{aligned} \tag{13}$$

is a $\mathbb{P}_\mu$ ($\mathcal{F}_\cdot$-) martingale with quadratic variation

$$\int_0^t \gamma\langle \phi^2, X_s^\theta\rangle\, ds,$$

as required. □

Notice now that if $r^2 h(r) \to \infty$ as $r \to \infty$, then h^θ grows without bound as $\theta \to \infty$. This suggests that the negative part of the drift in (13) will eventually dominate and the process will die out.

REMARK 3.3. If the spatial motion were a stable process of index β, then the rescaling

$$\langle \phi, X_t^\theta\rangle \triangleq \left\langle \frac{1}{\theta^\beta}\phi\left(\frac{x}{\theta}\right), X_{\theta^\beta t}(dx)\right\rangle,$$

would lead to a solution to the martingale problem with parameters $(h^\theta, M, \theta^\beta\alpha, \gamma)$ where

$$h^\theta(r) = \theta^\beta h(\theta r).$$

For interaction kernels that decay more quickly, this rescaling also leads to an heuristic argument that the process will survive (at least for some choices of the other parameters) provided we are *not* working in the delicate case of dimension two. To see this, and also to explain the restriction on the possible forms of singular interaction kernel, we recall some well-known facts about the Dawson–Watanabe superprocess. Recall that for a finite random measure X on $\mathbb{R}^d$, the associated *Campbell measure*, Q, on $\mathcal{M}_F(\mathbb{R}^d) \times \mathbb{R}^d$ is defined by

$$Q(A \times B) = \mathbb{E}[\chi_A(X)X(B)].$$

If the *mean measure*, I, is now defined by

$$I(B) = \mathbb{E}[\langle \chi_B, X\rangle],$$

then the *Palm distributions*, $\mathbb{P}^{(x)}$, are characterized by

$$\int_{\mathbb{R}^d}\int_{\mathcal{M}_F(\mathbb{R}^d)} g(x,\mu)\mathbb{P}^{(x)}(d\mu)I(dx) = \int_{\mathcal{M}_F(\mathbb{R}^d)} g(x,\mu)\mu(dx)\mathbb{P}(d\mu).$$

The intuitive idea is that if (X, Y) denote the coordinate variables in $\mathcal{M}_F(\mathbb{R}^d) \times \mathbb{R}^d$, then under the Campbell measure, Y is a point chosen at random according to X and $\mathbb{P}^{(x)}$ is the conditional distribution of X given that $Y = x$. For example,

$$\mathbb{P}^{(x)}[\langle \chi_{B(x,r)}, X\rangle],$$

tells us the expected amount of mass in a ball of radius r about the point x, conditional on x being a point in the support of X.

We define $\psi_d(r)$ by

$$\psi_d(r) \sim \begin{cases} r, & d = 1,\\ r^2\log(1/r), & d = 2,\\ r^2, & d \geq 3.\end{cases}$$

In the case of the (critical) Dawson–Watanabe superprocess [corresponding to $\alpha = 0$ in (8)] denoted by Z_t, it is an easy calculation to check that

$$\lim_{r\downarrow 0}\mathbb{P}^{(x)}\left[\frac{\langle \chi_{B(x,r)}, Z_t\rangle}{\psi_d(r)}\right] = k_d, \tag{14}$$

where the constant k_d is universal if $d \geq 2$, but may depend on x, t and Z_0 if $d = 1$ [see, e.g., Proposition 6.20 of Etheridge (2000)]. In other words, for "typical" r and x sampled from the support of the superprocess, the amount of mass in a ball of radius r about x is $\sim \psi_d(r)$ as $r \downarrow 0$.

In dimensions at least three, if x is a typical point of the support of the superprocess, then $\langle h^\theta(\|x-y\|), Z_t(dy)\rangle$ will be $O(1)$ as $\theta \to \infty$. In fact, the same is true in one dimension once one has chased through the dependence of k_1 on Z_0. This gives us hope that, at least for some parameters, our interacting process will survive. In two dimensions, however, the same quantity will grow like $\log\theta$ as $\theta \uparrow \infty$ and thus, in particular, dominate M. This suggests that for large θ, provided that the estimate (14) holds good, the process $\{X_t^\theta\}_{t\geq 0}$ can be compared to a *subcritical* superprocess (corresponding to $h = 0$, $\alpha < 0$) and, in particular, we cannot expect a nontrivial long-time limit. However, after our rescaling, in two dimensions, the estimate (14) is really a reflection of the large scale clumps in the superprocess model that our density dependent regulation is specifically designed to banish. Survival of the population will reflect successful eradication of clumping.

REMARK 3.4. As we remarked in the Introduction, it is possible to construct the process even for certain *singular* interaction kernels. The equation (14) determines the restrictions that we require on the singularity: $\langle h(\|x-y\|), Z_t(dy)\rangle$ should be finite for almost all x from the support of the superprocess Z. This requires $\int_{0+} h(r)\psi_d'(r)\,dr$ to be finite.

4. Applying the extinction proof to equation (6).

In this section we illustrate the proof of the first part of Theorem 1.5 by applying it to prove the following lemma.

LEMMA 4.1. *For any choice of* (*strictly*) *positive constants* α, M, λ *and* γ, *the solution* x_t *to equation* (6) *will hit zero in finite time.*

Our method is by no means the easiest way to prove this result, but it is useful in explaining the idea, and the limitations, of our method of proof of extinction.

The idea is simple. We think of $\{x_t\}_{t\geq 0}$ as being partitioned into distinct families. We then compare the evolution of the process to one in which fecundity is reduced, not in proportion to the whole population, but instead according to the size of an individual's own (extended) family. That we can make this comparison is a consequence of the following lemma.

LEMMA 4.2. *Suppose that* $x_0 = x_0^1 + x_0^2$ *for some* $x_0^1, x_0^2 \geq 0$. *Then we can construct a triple* $\{x_t, x_t^1, x_t^2\}_{t\geq 0}$ *such that the marginal processes*, $\{x_t\}_{t\geq 0}$, $\{x_t^1\}_{t\geq 0}$, $\{x_t^2\}_{t\geq 0}$, *are* (*weak*) *solutions to equation* (6) *with initial conditions* x_0, x_0^1, x_0^2, *respectively, and such that*

$$x_t \leq x_t^1 + x_t^2 \qquad \forall t \geq 0.$$

PROOF. In order to mimic our construction in the spatial setting as closely as possible, we proceed by constructing a (tight) sequence of processes, $\{\{x_t^n, x_t^{1,n}, x_t^{2,n}\}_{t\geq 0}\}_{n\geq 1}$, in such a way that

$$x_t^n \leq x_t^{1,n} + x_t^{2,n} \qquad \forall t \geq 0, n \geq 1,$$

and any limit point of the sequence has the desired marginal distributions.

First we choose a sequence of initial conditions in such a way that

$$(x_0^n, x_0^{1,n}, x_0^{2,n}) \to (x_0, x_0^1, x_0^2) \qquad \text{as } n \to \infty$$

and $nx_0^n, nx_0^{1,n}, nx_0^{2,n} \in \mathbb{N}$ for all $n \in \mathbb{N}$.

Exactly as for our construction of the process in Section 2, we start with a sequence of auxiliary processes, denoted $\{\{y_t^n\}_{t\geq 0}\}_{n\geq 1}$. Since there is no spatial component, y_t^n can be taken to be the *size* of the Y_t^n population described in Section 2, measured in units of size n. The limit as $n\to\infty$ will solve equation (6) with $\lambda = 0$ and $y_0 = x_0$. In other words the limit is a supercritical continuous state branching process, started from x_0. Once again we augment the process $\{y_t^n\}_{t\geq 0}$ by equipping each individual with an independent uniform [0, 1] random variable and a label.

Only the label will be more complex. It will take the form $\{i, j, k\}$, where $i \in \{1, 2\}$ and $j, k \in \{0, 1\}$. It is natural to choose $y_0^n = x_0^n$ for all n. We then assign the label i so that $nx_0^{i,n}$ individuals of the y_0^n-population have label i. This label will be passed down from mother to child and so simply refers to the type of the ancestor at time zero. The label j is 1 or 0 according to whether the individual forms part of $\{x_t^n\}$ or not, and k is 1 or 0 according to whether the individual is in $\{x_t^{i,n}\}$ or not. Note that all individuals in $x^{1,n}$ (resp. $x^{2,n}$) have $i = 1$ (resp. $i = 2$).

The assignment of the labels j and k is determined as follows. For all individuals $j = 1 = k$ at time zero. If a parent has $j = 0$, so do her descendants. If an individual with label $j = 1$ dies at time τ, then each of her descendants (in the tree determined by the y^n-population) has label $j = 0$ or $j = 1$ according as their uniform random variable U satisfies

$$U < \frac{\alpha\lambda x_{\tau-}^n}{n} \quad \text{or} \quad U > \frac{\alpha\lambda x_{\tau-}^n}{n}.$$

Finally we turn to the label k. If a parent has $k = 0$, so do her descendants. If a parent has label $(i, 0, 1)$ her descendants have label $(i, 0, 0)$ or $(i, 0, 1)$ according as

$$U < \frac{\alpha\lambda x_{\tau-}^{i,n}}{n} \quad \text{or} \quad U > \frac{\alpha\lambda x_{\tau-}^{i,n}}{n}.$$

The only slightly tricky case is if the parent has type $(i, 1, 1)$. We wish to arrange that no individual will ever have label $(i, 1, 0)$. The label of each descendant of a type $(i, 1, 1)$ individual is therefore determined by the following rules: if $U > \alpha\lambda x_{\tau-}^{i,n}/n$, then $k = 1$ and j is determined as above. If $U < \alpha\lambda x_{\tau-}^{i,n}/n$, then we consider two cases separately. If $U < \alpha\lambda x_{\tau-}^{n}/n$, then the offspring is of type $(i, 0, 0)$. If

$$\frac{\alpha\lambda x_{\tau-}^n}{n} < U < \frac{\alpha\lambda x_{\tau-}^{i,n}}{n},$$

then the offspring is of type $(i, 1, 1)$, but another individual, sampled at random from those of type $(i, 0, 1)$ has her type changed to $(i, 0, 0)$. [Note that we can always do this since if there does not exist an individual of type $(i, 0, 1)$, then all individuals from $x^{i,n}_{\tau-}$ have label $j = 1$ and so belong to $x^n_{\tau-}$, which implies $x^n_{\tau-} \geq x^{i,n}_{\tau-}$.]

Since the $x^{i,n}$-population is exchangeable, this process of removing an individual of type $(i, 0, 1)$ from the $x^{i,n}$-population to avoid the creation of an $(i, 1, 0)$ individual, retains the distribution of $x^{i,n}_t$ as that of continuous time branching in which a death at time τ leaves mean number of offspring

$$\left(1 + \frac{\alpha(M - \lambda x^{i,n}_{\tau-})}{n}\right)$$

and the variance of the number of offspring is $\sigma^2 + O(1/n)$.

With these definitions,

$$x^n_t \leq x^{1,n}_t + x^{2,n}_t.$$

Tightness is a triviality and evidently any limit point has the right marginal distribution and so the proof is complete. □

PROOF OF LEMMA 4.1. First recall [e.g., from Knight (1981)], that the *natural scale* of the diffusion governed by (6) is given by

$$s(x) = \int_1^x \exp\left(-\int_1^y \frac{2\alpha}{\gamma}(M - \lambda z)\, dz\right) dy.$$

Now for a given constant A and $0 < x < A$, the probability that the process hits A before it hits zero, given that it starts from x, is given by

$$\frac{s(x) - s(0)}{s(A) - s(0)} = \frac{\int_0^x \exp(-\int_0^y (2\alpha/\gamma)(M - \lambda z)\, dz)\, dy}{\int_0^A \exp(-\int_0^y (2\alpha/\gamma)(M - \lambda z)\, dz)\, dy}. \tag{15}$$

We now apply Lemma 4.2 iteratively. Suppose that $x_0 = \eta$. Then, for any $N \in \mathbb{N}$, the solution to equation (6) is dominated (in law) by the sum of N independent copies of the solution to (6) with initial value η/N. We shall refer to these solutions as *clusters*.

Each cluster is allowed to evolve until it hits either zero or A (with the constant A to be chosen). The expected time before it does one or the other is finite. If it hits A, then it is subdivided into N new clusters that form the next "generation." Evidently, the mean number of "offspring" of each cluster is

$$N \frac{s(\eta/N) - s(0)}{s(A) - s(0)},$$

and using equation (15), provided N is sufficiently large that $\eta\lambda/N < 2M$ this is bounded above by

$$\mu(\eta, M, \alpha) = \frac{\eta}{\int_0^A \exp(-\int_0^y (2\alpha/\gamma)(M - \lambda z)\, dz)\, dy}. \tag{16}$$

Using the strong Markov property, we now go through the same process for each offspring cluster. Thus each cluster has either zero or N offspring and provided that $A\lambda/N < 2M$, the expected number of offspring of a cluster is at most $\mu(A, M, \alpha)$.

Now we choose A sufficiently large that $\mu(A, M, \alpha)$ is less than one and N sufficiently large that $A\lambda/N < 2M$. The cluster process is then a subcritical (age dependent) branching process and so dies out in finite time. □

Evidently the key to this proof is that we can choose A to be as large as we please. The point is that if we consider our process as being made up of "family trees," then we are estimating the whole process from above by one in which individuals are not killed according to the whole population, but instead only according to their close relatives. If A is large, then the time until a cluster hits A is long, by when the number of surviving clusters is small. This means that killing according to close relatives at least constitutes killing according to a significant proportion of the population. For small A, the estimate is poor as killing only by close relatives is ignoring the bulk of the population.

5. Proof of Theorem 1.5 I: extinction. We now adapt the proof of the last section to the spatial setting. Again the basic idea is that we kill an individual in the population only according to its close relatives. However, the analogue of Lemma 4.2 requires an extra twist as a naive approach fails. In order to justify the relabelling of the individuals in $x_{\cdot}^{i,n}$, we exploited exchangeability. If h is nonconstant, then we lose that property. We therefore fix constants $\rho, \lambda > 0$ such that $\lambda = h(3\rho)$ [so $h(r) > \lambda$ for $r < 3\rho$] and dominate the size of each family tree by that of one in which local regulation is governed not by h, but by the constant λ times the family size. In other words, until the diameter of a "cluster" of the process is at least 6ρ, we dominate its mass by a solution to equation (6).

Again to formally construct the coupling, we work with a sequence of approximating particle systems. Suppose then that $\{X_0^n\}_{n\geq 1}$ is a sequence of (finite) purely atomic measures, all of whose atoms have mass $1/n$, such that $X_0^n \to X_0$ as $n \to \infty$. The first step is to subdivide the initial measure in such a way that each portion is contained in a ball of radius ρ. These portions play the rôle of x_0^1 and x_0^2 of Lemma 4.2 in providing "ancestors" of clusters. Suppose then that we are constructing the cluster descended from mass contained in the ball of radius ρ centred on some point x_0, denoted $B(x_0, \rho)$. Again we construct both the process $\{X_t^n\}_{t\geq 0}$ and the dominating process, that we now denote $\{\tilde{X}_t^n\}_{t\geq 0}$, as subpopulations of a (supercritical) branching process, $\{Y_t^n\}_{t\geq 0}$. The evolution of the Y^n-process is precisely that described in Section 2. We also exploit precisely the same device of attaching an independent uniform $[0, 1]$ random variable to each individual in the Y-population.

An individual of Y^n born of a parent in X^n at time τ and position x, will herself form part of X^n if her uniform random variable satisfies

$$U > \frac{\alpha}{n}\langle h(x, y), X_{\tau-}^n(dy)\rangle.$$

We define $\tilde{X}^n$ exactly as before, by dictating that a descendant of an individual in $\tilde{X}^n$ is also in $\tilde{X}^n$ if

$$U > \frac{\alpha}{n}\langle \lambda, \tilde{X}^n_{\tau-}\rangle$$

and moreover, as in Lemma 4.2, if necessary we perform a relabelling when a death occurs so that it is always an individual that is not a part of X^n that dies. Note that this relabelling does not alter the distribution of $|\tilde{X}^n|$. The cluster is stopped when either $|\tilde{X}^n| = A$ or 0 or Y^n hits the boundary, $\partial B(x_0, 3\rho)$, of the ball $B(x_0, 3\rho)$, but note that the Y and $\tilde{X}$-processes can be defined *for all time*.

Again we omit the proof of tightness of the approximating processes and passage to the limit and move straight on to using the coupled (limit) processes to prove our result.

We begin by proving part 1 of Theorem 1.5. Our second extinction result, part 2(a), will require more work.

In the same way as in Section 4, we must estimate the number of clusters for which $|\tilde{X}| = A$ or Y hits $\partial B(x_0, 3\rho)$ before $|\tilde{X}| = 0$. Evidently this is bounded by the sum of the number of clusters of the Y-process that *ever* hit $\partial B(x_0, 3\rho)$, plus the number of clusters of the $\tilde{X}$-process that reach mass A before they hit zero. (This is where we require that the Y and $\tilde{X}$ processes exist for all time.)

Consider first a single cluster that starts from mass η, contained in the ball $B(x_0, \rho)$ at time zero. We assess the probability that the corresponding cluster of the Y-process ever hits $\partial B(x_0, 3\rho)$. Now the supercritical superprocess started from η can itself be thought of as a Poisson number of independent clusters, with a parameter that depends on time, but which is finite at any positive time. Moreover, its support propagates at finite speed, so there exists $p(\rho, \alpha M) > 0$ such that the probability that the support ever hits $\partial B(x_0, 3\rho)$ is

$$1 - \exp\bigl(-p(\rho, \alpha M)\eta\bigr).$$

Evidently $p(\rho, \alpha M)$ is monotone increasing in αM. [These standard results can be found in our references on superprocesses. In order to prove 2(a) we will be much more precise about the rate of spread of the Y-clusters.]

We now proceed as before, partitioning each surviving cluster and estimating the number of these new clusters that survive to contribute to the next generation. We are implicitly using the strong Markov property. This is valid as our construction of the process enables us to compare the process to the sum of a countable number of copies corresponding to a partition of the initial condition into finite pieces. After such a partition, we are considering the process started only from finite initial conditions and for that process uniqueness of the solution to the martingale problem guarantees the strong Markov property.

Suppose then that a cluster contributes to the next generation, that is, either the Y-process hits $\partial B(x_0, 3\rho)$ or $|\tilde{X}| = A$ before $|\tilde{X}| = 0$. When one of these events happens, the total mass of the cluster is at most A and its diameter is at most 6ρ and so we can partition it into at most $d^{d/2} \cdot 3^d N$ pieces (offspring), each contained in a ball of radius ρ and each of mass at most A/N. The expected number of offspring clusters is then *at most*

$$d^{d/2} \cdot 3^d N(1 - \exp(-p(\rho, \alpha M)A/N)) + d^{d/2} \cdot 3^d \mu(A, M, \alpha),$$

where μ is defined in (16). We wish to be able to choose this to be less than one. We can no longer choose A to be arbitrarily large as that increases the first term which, for large N, is about $d^{d/2} \cdot 3^d A p(\rho, \alpha M)$. Instead, we choose A small enough that $d^{d/2} \cdot 3^d A p(\rho, \alpha M) < 1/2$ and then, with A fixed, we choose α_0 large enough that for $\alpha > \alpha_0$, $d^{d/2} \cdot 3^d \mu(A, K/\alpha, \alpha) < 1/2$. By comparison with a subcritical spatial branching process, the proof of extinction for $\alpha > \alpha_0$ is complete.

To justify 2(a) we must work a little harder. We now fix $\alpha > 0$. As before we consider clusters of the process initially contained in a ball of radius ρ and follow their evolution until (at the very latest) the time when they first hit the exterior of a concentric ball of radius 3ρ. We will later take ρ to be large. The process $|\tilde{X}|$ is governed by equation (6) with $\lambda = h(3\rho)$. We may assume without loss of generality that $h(3\rho) < 1$ as by assumption (see Notation 1.2) $h(\theta) \to 0$ as $\theta \to \infty$. [For the case of finite initial measures, if $h(\theta) \nrightarrow 0$ as $\theta \to \infty$, we can compare the whole process to the solution to equation (6).] We also choose A such that

$$\frac{1}{A} \int_0^A \exp\left(-\int_0^y \frac{2\alpha}{\gamma}(1 - z)\, dz\right) dy > 2 \cdot d^{d/2} \cdot 3^d.$$

Notice then that for $\lambda, M < 1$,

$$\begin{aligned} &\frac{\lambda}{A} \int_0^{A/\lambda} \exp\left(-\int_0^y \frac{2\alpha}{\gamma}(M - \lambda z)\, dz\right) dy \\ &\quad = \frac{1}{A} \int_0^A \exp\left(-\int_0^y \frac{2\alpha}{\lambda\gamma}(M - z)\, dz\right) dy \geq 2^{1/\lambda} \cdot d^{d/2\lambda} \cdot 3^{d/\lambda}. \end{aligned} \tag{17}$$

This guarantees that the probability that $|\tilde{X}|$, started from η/λ hits A/λ before it hits zero is at most $\eta/(2 \cdot d^{d/2} \cdot 3^d)^{1/\lambda}$.

We now consider the probability that the Y-cluster has reached the boundary of the ball of radius 3ρ by time t. First observe that by a trivial modification of the work of Iscoe (1986), for any positive test function, ψ,

$$\mathbb{E}\left[\exp\left(-\int_0^t \langle \psi, Y_s \rangle\, ds\right)\right] = \exp(-\langle u(t, x), X_0(dx) \rangle), \tag{18}$$

where

$$\frac{\partial u}{\partial t}(t, x) = \frac{1}{2}\Delta u(t, x) + \alpha M u(t, x) - \frac{\gamma}{2} u^2(t, x) + \psi(x), \qquad u(0, x) = 0.$$

In order to estimate the probability that the Y-cluster has not reached the boundary of the ball of radius 3ρ by time t, we take ψ in equation (18) to be k times the indicator function of the complement of that ball and let $k \to \infty$. The resulting sequence of functions $u_k(t,x)$ is monotone increasing in k to a finite limit. This can be proved exactly as in Dawson, Iscoe and Perkins (1989). Indeed using Lemma 3.5 of that paper we can say much more. They are concerned with the case $\alpha M = 0$, but a simple comparison shows that

$$\lim_{k\to\infty} u_k(t,x) \le e^{\alpha M t} \lim_{k\to\infty} u_0(t,x). \tag{19}$$

Dawson, Iscoe and Perkins estimate, in the case $\alpha M = 0$, the probability that a Y-cluster centered on a point distance R from the origin at time zero charges a unit ball about the origin by time t

$$1 - \exp\left(-\frac{C}{R^2} P(t,R)\langle 1, Y_0\rangle\right),$$

where $P(t,R)$ is the probability that a Brownian motion started from a point at distance R from the origin at time zero has hit the unit ball centered at the origin by time t. The constant C depends only on γ. Exploiting the observation (19) and by covering the boundary of the ball of radius 3ρ by $O(\rho^{d-1})$ unit balls we see that the probability that a cluster of our Y-process of initial mass $\eta/h(3\rho)$ contained in a ball of radius ρ has hit the boundary of the concentric ball of radius 3ρ by time t is certainly at most

$$C'\rho^{d-1}\left(1 - \exp\left(-\frac{C\eta}{\rho^2 h(3\rho)}\frac{1}{(2\pi t)^{d/2}} e^{\alpha M t - \rho^2/2t}\right)\right),$$

for a constant C' independent of the parameters of our model.

Finally we consider the probability that a cluster of $\tilde{X}$, started from mass $\eta/h(3\rho)$ hits zero before it hits mass $A/h(3\rho)$, but not by time t. This is evidently bounded by the probability that the total mass of a cluster of the Y process (with no conditioning) started from initial mass $\eta/h(3\rho)$ has not hit zero by time t which in turn is easily calculated to be

$$1 - \exp\left(-\frac{\alpha M \eta}{h(3\rho)(1 - e^{-\alpha M t})}\right).$$

We now combine these estimates. We subdivide our initial population into clusters of mass at most $\eta/h(3\rho)$ (where $\eta = A/N$) and each contained in a ball of radius ρ. For one such cluster, we allow the cluster to evolve until either the Y-process hits the boundary of the concentric ball of radius 3ρ *or* the $|\tilde{X}|$-process hits 0 or $A/h(3\rho)$ *or* time t, whichever happens first. When such an event happens, the total mass of the cluster will be at most $A/h(3\rho)$ and it will be subdivided into at most $3^d N = 3^d A/\eta$ "offspring" clusters, each with initial mass at most $\eta/h(3\rho)$ and contained in a ball of radius ρ. The probability that a cluster has descendants in the next generation is at most the probability that $|\tilde{X}|$ started from $\eta/h(3\rho)$ hits $A/h(3\rho)$ before zero, plus the probability that a cluster of the Y-process with initial mass $\eta/h(3\rho)$ has not hit zero by time t, plus the probability that a cluster of the Y-process hits the boundary of the ball of radius 3ρ by time t. Combining (17) with $\lambda = h(3\rho)$, with our calculations above, under the assumption that $M < 1$ and $h(3\rho) < 1$, the expected number of offspring clusters is therefore bounded above by

$$\begin{aligned}
&d^{d/2}\cdot 3^d\mu(A,M,\alpha) + \frac{d^{d/2}\cdot 3^d A}{\eta}\left(1-\exp\left(-\frac{\alpha M\eta}{h(3\rho)(1-e^{-\alpha Mt})}\right)\right)\\
&\quad + \frac{d^{d/2}\cdot 3^d AC'\rho^{d-1}}{\eta}\left(1-\exp\left(-\frac{C\eta}{\rho^2 h(3\rho)}\frac{1}{(2\pi t)^{d/2}}e^{\alpha Mt-\rho^2/2t}\right)\right)\\
&\quad \approx d^{d/2}\cdot 3^d\mu(A,M,\alpha) + \frac{d^{d/2}\cdot 3^d A\alpha M}{h(3\rho)(1-e^{-\alpha Mt})}\\
&\qquad + \frac{d^{d/2}\cdot 3^d ACC'\rho^{d-1}}{\rho^2 h(3\rho)}\frac{1}{(2\pi t)^{d/2}}e^{\alpha Mt-\rho^2/2t}.
\end{aligned}$$

Now set $M = h(3\rho)/(4\alpha A)$ and $t = 1/(\alpha M)$. Then the mean number of offspring is at most

$$d^{d/2}\cdot 3^d\mu(A,M,\alpha) + \frac{1}{4} + \frac{d^{d/2}\cdot 3^d ACC'\rho^{d-3}}{h(3\rho)}\frac{h(3\rho)^{d/2}}{(8\pi A)^{d/2}}e^{1-\rho^2 h(3\rho)/(8A)}.$$

Now since $\rho^2 h(\rho) > \varepsilon\rho^\delta$ for sufficiently large ρ, we can choose ρ so that the last term is at most $\frac{1}{4}$ and $M_0 \triangleq h(3\rho)/(4\alpha A) < 1$. We now see that the mean number of "offspring" of each cluster is then less than one and, once again comparing with a subcritical branching process, the result now follows.

REMARK 5.1. We can considerably improve the bounds in the proof above. For example we can estimate the rate of growth of a cluster of the X-process directly, rather than via that of a cluster of the Y-process. The key quantity is the total mass of the cluster up to time t. However, even with this extra work the method does not appear to be powerful enough to prove the result that we believe to be true, namely that $M_0 = \infty$ in the statement of Theorem 1.5.2(a).

6. Proof of Theorem 1.5 II: survival. In this section we sketch another comparison that shows that for certain parameters the process actually survives. Not surprisingly, in view of the work of Mueller and Tribe (1994), we compare our process to an *oriented percolation process* on $\mathbb{Z}_+ \times \mathbb{Z}^d$. The points $(n_1, \underline{x}_1)$ and $(n_2, \underline{x}_2)$ are neighbors in this lattice if $n_2 = n_1 + 1$ and $\underline{x}_1$ and $\underline{x}_2$ are neighbors in $\mathbb{Z}^d$. We think of the first component as encoding time and this determines the directions on the bonds in the lattice. Since we are thinking of this as a discrete time version of the contact process, we shall use the terms "infected" and "vacant" to indicate whether the population in a site is above or below a certain threshold.

Although the proof can be (somewhat messily) adapted, it is simplest to describe in the case of an already subdivided population. The idea is that we shall declare a deme to be *infected* (or occupied) if the population there has mass more than M/K, where

$$\frac{\alpha M}{K} = \frac{c}{2}.$$

Otherwise it is vacant. In order to make a comparison with an oriented percolation process, we wish to obtain a lower bound on the probability that a vacant site in $\mathbb{Z}^d$ becomes infected in the next discrete time interval and an upper bound on the probability that an infected site recovers. If for some parameters the probability of infection is high enough and the probability of recovery is low enough then we have percolation and the population persists.

Let us think for a moment in terms of the continuous time process. Infection arises from mass migrating from neighboring sites. (The notion of neighbor here is determined by the migration matrix.) Of course for the site to actually become infected, we have to wait until the total mass there is at least M/K. Now

$$\begin{aligned} dX_t(i) &= \sum_j m_{ij}\big(X_t(j) - X_t(i)\big)\,dt \\ &\quad + \alpha\left(M - \sum_j \lambda_{ij} X_t(j)\right) X_t(i)\,dt + \sqrt{\gamma X_t(i)}\,dB_t^{(i)}. \end{aligned}$$

If the site i is not infected, then $X_t(i) < M/K$ and so (after some rearrangement) we see that

$$\begin{aligned} dX_t(i) &\geq \sum_j (m_{ij} - \alpha\lambda_{ij} M/K) X_t(j)\,dt \\ &\quad + \left(\alpha M - \sum_j m_{ij}\right) X_t(i)\,dt + \sqrt{\gamma X_t(i)}\,dB_t^{(i)}. \end{aligned}$$

Note that if a site j is infected, then $X_t(j) \geq M/K$. Thus, choosing K (e.g., as above) in such a way that $m_{ij} > 2\alpha\lambda_{ij} M/K$ and M large enough that $M\alpha > \Sigma_j m_{ij}$, we see that mass immigrates into the uninfected deme at a rate that is bounded below by a constant multiple of the number of infected neighbors, and once there it evolves according to a supercritical continuous state branching process. Evidently by choosing M large, we can make the time to infection arbitrarily small.

Now consider the recovery rate. Recovery arises because of crowding of a region or because of fluctuations caused by the Feller noise. However, in a crowded region, the negative effect on fecundity may be offset by increased immigration from crowded neighbors. Suppose then that we are at an occupied site. We consider the recovery rate only when the total mass is less than $2M/K$. Since the total mass process is continuous, it will be enough to bound this rate above (or equivalently the growth rate below) as excursions above level $2M/K$ simply add to the total time to recovery. Then at the infected site,

$$dX_t(i) \geq \sum_j \left(m_{ij} - 2\alpha\lambda_{ij}\frac{M}{K}\right)X_t(j)\,dt$$
$$+ \left(\alpha M - \sum_j m_{ij}\right)X_t(i)\,dt + \sqrt{\gamma X_t(i)}\,dB_t^{(i)}.$$

Since by our choice of K, $m_{ij} > \alpha\lambda_{ij}2M/K$, provided that, as above, $\alpha M > \sum_j m_{ij}$, we see that this is a supercritical continuous state branching process (with growth rate $\alpha M - \sum_j m_{ij}$) with an additional positive drift.

We should like to use these estimates to compare the process with a contact process, but since neither infection nor recovery rates are exponentially distributed, and since the infection of a site is subject to a "delay" (while the supercritical branching increases immigrated mass to M/K), it is simplest to make a comparison with a *discrete time* version of the contact process, namely oriented percolation. There are two difficulties. One is that infected sites may not remain infected during the whole discrete time period. The other is that the probability of being vacant or infected at different sites is *not* independent. This dependence forces comparison with 3-dependent oriented percolation. For full details of this process and how to make comparisons of this type, we refer to Durrett (1995).

The population certainly survives if the set of infected sites percolates. We now use our observations above for the continuous time process to make the necessary estimates for comparison with the oriented percolation. This final step is elementary. The probability that an infected site has recovered by time 1 is bounded above by the probability that a supercritical continuous state branching process with growth rate $\alpha M - \sum_j m_{ij}$ and a reflecting boundary at c/α has mass less than $c/(2\alpha)$ at time one, given that its initial mass is $c/(2\alpha)$. This quantity decreases monotonically to zero as $M \uparrow \infty$. On the other hand, the probability that a vacant site becomes infected is *at least* that resulting from all infected neighbors trying to produce an infection. In particular, if it has at least one infected neighbor (and these are the only sites we care about), the probability of successful infection is in turn bounded below by the probability that the supercritical branching process corresponding to that infected neighbor (with growth rate $\alpha M - \Sigma_j m_{ij}$ and reflection at c/α), started from mass (at least) $c/(2\alpha)$ at time zero, never has mass less than $c/(4\alpha)$ in the time interval [0, 1] times the probability that a supercritical continuous state branching process with reflection at c/α, immigration at constant rate

$$\frac{c}{4\alpha}\left(m_{ij} - \lambda_{ij}\frac{c}{2}\right)$$

and growth rate

$$\alpha M - \sum_j m_{ij}$$

is at least $c/(2\alpha)$ at time 1. This is monotone increasing as $M \uparrow \infty$ to limit one. (Notice that this is where the dependence comes in. Whether or not this site is infected in the next generation is dependent on whether the "infecting" site remains infected or not.) Combining these observations, we see that for sufficiently large M the set of infected sites percolates, and so in particular our population survives for all time, with strictly positive probability.

REMARK 6.1. Formally, we should have constructed a coupling of the stepping-stone version of the Bolker–Pacala model and the oriented percolation process. We hope that the reader is convinced by our work in previous sections that this is possible via a passage to the limit.

7. Conclusions and discussion.

Our result is incomplete in a number of ways, but combined with our heuristics it does highlight the essential rôle of the spatial structure of the population in its longtime survival. We require the population to spread sufficiently quickly that families are able to colonize vacant neighboring sites before they are killed by overcrowding (caused by their own relatives). If the spatial spread is too slow, then the process behaves essentially as one divided into independent colonies.

As well as restricting our proof to special cases of the model, we introduced the caveat that for infinite initial measures our proof of extinction of the superprocess version of the model is valid only for solutions to the martingale problem obtained by the limiting procedure described in Section 2. Of course, we can argue that this is not really a restriction, since our motivation for studying the model is precisely as an approximation to such discrete branching particle systems. However, mathematically, the obstruction is that we have not proved uniqueness of the solution to the BP-martingale problem for processes taking values among tempered measures. It is almost certainly an adaptation of the ideas of Evans and Perkins (1998) to prove uniqueness of the martingale problem and also to construct the processes $\{\tilde{X}_t\}_{t\geq 0}$ that we use in our comparison lemma of Section 5 directly, without reversion to a passage to the limit.

The Bolker–Pacala models represent a very special class of density dependent population models. A natural extension to the model replaces the "linear" interaction term

$$a(x,\mu) = \big(M - \langle h(x,y), \mu(dy)\rangle\big),$$

by a nonlinear interaction

$$\tilde{a}(x,\mu) = f\big(a(x,\mu)\big),$$

for a suitably chosen function $f:\mathbb{R}\to\mathbb{R}$. In order to apply Dawson's Girsanov theorem to construct the process, we require some conditions on f. For example, it suffices that

$$\sup\{f(a(x,\mu)):\mu\in\mathcal{M}_F(\mathbb{R}^d)\}<\infty$$

and

$$\sup\{|f(a(x,\mu))|:\mu\in\mathcal{M}_F(\mathbb{R}^d),\,\langle 1,\mu\rangle\le n\}<\infty\qquad\forall n\in\mathbb{N}.$$

Our extinction proof will go through with only minor changes. Our proof of the survival of the corresponding stepping-stone model did exploit linearity, but could certainly be adapted if, for example, $-f$ were sublinear.

When one observes a biological population, one is necessarily looking at a *conditional* distribution. The very fact that the population is there means that in effect we are conditioning on nonextinction. It is not unnatural then to condition our stochastic model on nonextinction. In the branching process setting, the model is still inadequate as it predicts that surviving populations should grow without bound (linearly in time for critical branching processes and exponentially in time for noncritical processes). For equation (6) one can condition on nonextinction and the conditioned process has finite moments of all orders (for all time). In the spatial setting, however, it is far from clear how to proceed with such a conditioning. This highlights one of the problems with the Bolker–Pacala model. It seems to be extremely difficult to find analytic expressions even for the simplest quantities.

Our original motivation for studying the Bolker–Pacala models was an interest in population genetics. Before modelling the genetic evolution of a spatially structured population, one needs an ecological model. Armed with this one then hopes to find expressions for observable quantities such as correlations in allele frequencies that might then allow us to actually *measure* forces such as selection. The fact that, at least for some choices of parameter, the Bolker–Pacala model is governed by local rules and yet predicts stable population dynamics is undoubtedly appealing. However, there are some difficulties with implementing it. As we remarked in Section 2, one of the attractions of the Feller diffusion is that it approximates a wide variety of local structures by a model that is characterized by just two parameters. Moreover, these parameters can be expressed in terms of measurable characteristics of the population (the mean and variance of the number of offspring of an individual plus the mean inter-generation time). The Dawson–Watanabe superprocess retains this feature, requiring just one extra parameter (a measure of dispersal) in its specification. Unfortunately, unless one postulates that h belongs to a parametrized class of functions, in the Bolker–Pacala setting we lose this property. Nonetheless, we hope that over sufficiently large spatial scales, the detailed structure of h should not be important. It would certainly be interesting to know the implications of using this model in place of, say, a stepping stone model in population genetics. Such calculations appear to be highly nontrivial.

In summary, although the Bolker–Pacala model has some appealing features, what we should like is a class of models that approximate a wide variety of local structures (including those implied by the Bolker–Pacala model) and are parametrized in terms of a *small* number of *measurable* parameters. We shall return to this issue in future work.

Acknowledgments. This work arose out of a joint project with Professor Nick Barton of the Institute of Cell, Animal and Population Biology in the University of Edinburgh and I am grateful to him for useful discussions. For more details of the biological motivation for studying models of locally regulated populations, see, for example, Barton, Depaulis and Etheridge (2002) and references therein. I also thank Anton Wakolbinger and an anonymous referee for some very helpful comments.

REFERENCES

BARTON, N. H., DEPAULIS, F. and ETHERIDGE, A. M. (2002). Neutral evolution in spatially continuous populations. *Theoretical Population Biology* **61** 31–48.

BOLKER, B. M. and PACALA, S. W. (1997). Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theoretical Population Biology* **52** 179–197.

BOLKER, B. M. and PACALA, S. W. (1999). Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal. *American Naturalist* **153** 575–602.

DAWSON, D. A. (1993). Measure-valued Markov processes. *Ecole d'été de probabilités de Saint Flour XXI. Lecture Notes in Math.* **1541** 1–260. Springer, New York.

DAWSON, D. A., ISCOE, I. and PERKINS, E. A. (1989). Super-Brownian motion: Path properties and hitting probabilities. *Probab. Theory Related Fields* **83** 135–205.

DURRETT, R. (1995). Ten lectures on particle systems. *Ecole d'été de probabilités de Saint Flour XXIII. Lecture Notes in Math.* **1608** 97–201. Springer, New York.

DURRETT, R. and PERKINS, E. A. (1999). Rescaled contact processes converge to super-Brownian motion in two or more dimensions. *Probab. Theory Related Fields* **114** 309–399.

ETHERIDGE, A. M. (2000). *An Introduction to Superprocesses* **20**. AMS, Providence, RI.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. Wiley, New York.

EVANS, S. N. and PERKINS, E. A. (1994). Measure-valued branching diffusions with singular interactions. *Canad. J. Math.* **46** 120–168.

EVANS, S. N. and PERKINS, E. A. (1998). Collision local times, historical stochastic calculus and competing species. *Electron. J. Probab.* **3** 1–120.

FELLER, W. (1951). Diffusion processes in genetics. *Proc. Second Berkeley Symp. Math. Statist. Probab.* 227–246. Univ. California Press, Berkeley.

FELSENSTEIN, J. (1975). A pain in the torus: Some difficulties with the model of isolation by distance. *American Naturalist* **109** 359–368.

ISCOE, I. (1986). A weighted occupation time for a class of measure-valued critical branching Brownian motions. *Probab. Theory Related Fields* **71** 85–116.

KIMURA, M. (1953). Stepping stone model of population. *Ann. Rep. Nat. Inst. Genetics Japan* **3** 62–63.

KNIGHT, F. B. (1981). *Essentials of Brownian Motion and Diffusion*. Amer. Math. Soc., Providence, RI.

LAW, R., MURRELL, D. J. and DIECKMANN, U. (2003). On population growth in space and time: Spatial logistic equations. *Ecology* **84** 252–262.

LE GALL, J.-F. (1999). *Spatial Branching Processes, Random Snakes and Partial Differential Equations*. Birkhäuser, Basel.

MUELLER, C. and TRIBE, R. (1994). A phase-transition for a stochastic pde related to the contact process. *Probab. Theory Related Fields* **100** 131–156.

PERKINS, E. A. (2002). Dawson–Watanabe superprocesses and measure-valued diffusions. *Ecole d'été de probabilités de Saint Flour. Lecture Notes in Math.* **1781** 125–329. Springer, New York.

REVUZ, D. and YOR, M. (1999). *Continuous Martingales and Brownian Motion*. Springer, New York.

WALSH, J. B. (1986). An introduction to stochastic partial differential equations. *Ecole d'été de probabilités de Saint Flour. Lecture Notes in Math.* **1180** 265–439. Springer, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
1 SOUTH PARKS ROAD
OXFORD OX1 2JG
UNITED KINGDOM
E-MAIL: alison.etheridge@magd.ox.ac.uk
URL: www.stats.ox.ac.uk/~etheridg