

Probabilistic Foundations of Spatial Mean-Field Models in Ecology and Applications*

Denis D. Patterson[†], Simon A. Levin[‡], Carla Staver[§], and Jonathan D. Touboul[¶]

Abstract. Deterministic models of vegetation often summarize, at a macroscopic scale, a multitude of intrinsically random events occurring at a microscopic scale. We bridge the gap between these scales by demonstrating convergence to a mean-field limit for a general class of stochastic models representing each individual ecological event in the limit of large system size. The proof relies on classical stochastic coupling techniques that we generalize to cover spatially extended interactions. The mean-field limit is a spatially extended non-Markovian process characterized by nonlocal integro-differential equations describing the evolution of the probability for a patch of land to be in a given state (the generalized Kolmogorov equations (GKEs) of the process). We thus provide an accessible general framework for spatially extending many classical finite-state models from ecology and population dynamics. We demonstrate the practical effectiveness of our approach through a detailed comparison of our limiting spatial model and the finite-size version of a specific savanna-forest model, the so-called Staver–Levin model. There is remarkable dynamic consistency between the GKEs and the finite-size system in spite of almost sure forest extinction in the finite-size system. To resolve this apparent paradox, we show that the extinction rate drops sharply when nontrivial equilibria emerge in the GKEs, and that the finite-size system’s quasi-stationary distribution (stationary distribution conditional on nonextinction) closely matches the bifurcation diagram of the GKEs. Furthermore, the limit process can support periodic oscillations of the probability distribution and thus provides an elementary example of a jump process that does not converge to a stationary distribution. In spatially extended settings, environmental heterogeneity can lead to waves of invasion and front-pinning phenomena.

Key words. interacting particle system, Markov jump processes, integro-differential equations, ecological modeling, mean-field limit, spatial dynamics

AMS subject classifications. 60F05, 37N25, 34K60, 45K05

DOI. 10.1137/19M1298329

1. Introduction. Bridging the microscopic and macroscopic scales in theoretical ecology is an important challenge. Detailed microscopic models describing single organisms are

*Received by the editors November 7, 2019; accepted for publication (in revised form) by J. Sieber August 28, 2020; published electronically December 8, 2020.

<https://doi.org/10.1137/19M1298329>

Funding: The work of the second author was partially supported by NSF DMS 1615585 and 1951358. The work of the third author was partially supported by NSF DMS 1615531 and 1951394. The work of the fourth author was partially supported by NSF DMS 1951369.

[†]Department of Mathematics, Brandeis University, Waltham, MA 02453 USA (denispatterson@brandeis.edu).

[‡]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544 USA (slevin@princeton.edu).

[§]Department of Ecology and Evolutionary Biology, Yale University, Osborne Memorial Labs, New Haven, CT 06520 USA (carla.staver@yale.edu).

[¶]Department of Mathematics and Volen National Center for Complex Systems, Brandeis University, Waltham, MA 02453 USA (jtouboul@brandeis.edu).

extremely useful for their realistic interpretations and quantitative match to data, but their complexity often precludes a detailed mathematical analysis and obscures the key mechanisms at play [2, 25]. Simple models, on the other hand, allow for an in-depth mathematical understanding but necessitate simplifications and hypotheses that limit their predictive ability [49]. Despite these limitations, simple systems based on ordinary differential equations (ODEs) have proven extremely useful in ecology.

We propose bridging the gap between ecological models with macroscopic viewpoints and microscopic descriptions of stochastic transitions. While the methods presented here are quite general, for definiteness we apply them to a vegetation ecosystem and the Staver–Levin model of tropical vegetation cover [47, 49, 56]. The Staver–Levin model is a system of nonlinear differential equations describing the evolution of the fraction of landscape covered by grass, savanna trees, or forest trees. The two tree species differ in their birth and death rates as well as in the way they are affected by fires. Fires, carried by grass, kill (or burn) forest trees but not savanna trees; adult savanna trees resist fires, while saplings, *top-killed* by fires (i.e., burned but with the ability to resprout later), have their maturation into adult trees delayed. Each species also reproduces and dies with specific rates. This model displays complex dynamical behaviors, including multistability, limit cycles, and homoclinic and heteroclinic orbits. These complex dynamical structures are affected by the presence of noise, which possibly leads to stochastic resonances for Brownian perturbations [56]. From the ecological viewpoint, the Staver–Levin model summarizes, at a macroscopic scale, a variety of “microscopic” events arising at different spatial locations: seed dispersal leading to the birth of a new tree, growth of a new tree, occurrence of fires, death, etc. These events are intrinsically random and localized and depend on local interactions (e.g., fire spread or seed dispersal). A model explicitly accounting for these stochastic dynamics is essential to better understand how randomness affects the behavior of these systems and how dynamical behaviors are impacted by spatial interaction; these are two particularly important issues with regard to understanding ecosystem persistence and the potential impact of climate change on tropical vegetation distributions.

There is an extensive literature regarding limits of interacting particle systems dating back to the early works of Bernoulli [3] and Clausius [10] and to the celebrated work of Boltzmann [5] on the kinetic theory of gases. These works, aimed at relating the movement of molecules in a gas to macroscopic quantities, such as pressure, have seen remarkable development in recent years and have been applied to a variety of models (see, e.g., [7, 12, 13, 21, 40, 50]). The aforementioned mathematical framework has been widely used to derive and analyze spatially extended models in neuroscience [37, 44, 53, 54] and, to a lesser extent, in ecology [18]. The recent works of Durrett and Ma [20] and Durrett and Zhang [22] are important contributions and are most closely related to the present work. These authors consider a two state version of the Staver–Levin model on a toroidal lattice, and by taking an appropriately scaled spatial limit, they show convergence in probability of their interacting particle system to the solution of an integro-differential equation (IDE). They obtain a coexistence result by analyzing the resulting IDE and provide bounds on the coexistence time in terms of the system size. We adopt an alternative approach based on stochastic coupling methods [17, 38, 50, 53] that allows us to directly demonstrate convergence of the spatial Markov jump process to a McKean–Vlasov jump process (i.e., a process whose transition rates depend on the law of the process itself). Our framework incorporates nonlocal interactions between a finite number of

species and heterogeneity from random initial particle placement, and it can be applied to many other models in ecology and population dynamics.

We suppose that the locations of the vegetation, which we refer to as sites, are randomly distributed on the domain according to some probability measure q , and we consider a scaling limit in which sites are successively added to the domain according to this measure. This general model covers two cases relevant for applications: (i) *mesoscale models* which resolve the fine spatial structures of the interactions while still taking into account collective effects (q is a continuous distribution), and (ii) *macroscale models* which lose the fine spatial organization of the interaction between sites in favor of collective stereotyped interactions between a number of isolated vegetation populations (q is a finite combination of Dirac masses). In this way, the macroscale framework is reminiscent of a metapopulation or network model.

Within a unified framework that encapsulates the Stayer–Levin model and its variants, we show convergence of a class of spatial particle systems to their *mean-field limit*, a non-Markovian stochastic process. The mean-field limiting process will be characterized by IDEs on the probability density of the process that will be referred to as *generalized Kolmogorov equations* (GKEs). Similar to classical Kolmogorov equations for Markov processes, these equations govern the evolution of the probability distribution of the process. However, in contrast with classical Kolmogorov equations, the GKEs associated with the mean-field limit are nonlinear and, in the mesoscale model, nonlocal. Our spatial extensions of the Stayer–Levin model permit analysis of qualitative properties of practical interest which are not easily accessible via probabilistic methods. For instance, techniques developed for studying pattern formation [6, 26, 28, 34], wave speeds, and invasion phenomena [32, 52] and responses to heterogeneous environments [27, 35] in spatially extended ecological models are readily applicable to the GKEs. We demonstrate waves of invasion numerically in the mesoscale model (see Figure 5A1–A6), and this model can also naturally incorporate environmental heterogeneity via the choice of the initial site distribution (see Figure 5B1–B3). In fact, in the mesoscale framework, we show that both our particle system and the GKEs of the corresponding mean-field limit can exhibit front pinning, a phenomenon of much practical interest that has previously been found in PDE-based ecological models [27, 57].

From an ecological standpoint, further analysis of the models derived in this paper will enable a more thorough understanding of the determinants of the savanna-forest boundary, particularly in the presence of precipitation gradients, resource limitations, and climate change [35, 59]. Beyond savanna models, the techniques developed here can further be applied to various models. In biology, these techniques could naturally allow addressing stochastic and nonlocal effects arising in other vegetation models (e.g., generic vegetation models [33]), models of infectious diseases [8], and embryonic development [42] where differentiation of cells depends on nonlocal signals and gradients. Other perspectives include applications to social models and variations of voter models [11] with spatial dimension. In the domain of infectious diseases, our framework could be used in a study of disease transmission that takes into account the role of spatial positions and nonlocal transmission in a spatially extended SIR-type model [36]. In that regard, it would be of interest to extend our results to individuals moving in space and determine the impact of motility on viral spread, which are two particularly timely questions with implications for public health policies.

The paper is organized as follows. In section 2.1 we motivate and define a stochastic

Staver–Levin model based on microscale events with nonlocal spatial interactions (referred to as the *finite-size stochastic Staver–Levin model* hereafter). We then outline our main mathematical result, Theorem 2.2.1, in section 2.2; this result shows convergence to a well-defined mean-field limiting process for a general class of spatial particle systems and covers the finite-size Staver–Levin model as a special case. In section 3 we compare the qualitative behavior of the finite-size stochastic Staver–Levin model with the GKEs of the corresponding mean-field limits in a series of numerical experiments. We explain long-term transient behaviors via quasi-stationary distributions (QSDs), demonstrate limit cycles in both spatial and nonspatial versions of our models, and show waves of invasion and front pinning. The proof of Theorem 2.2.1 and supporting mathematical preliminaries are deferred to section 4. Further details of numerical parameters and routines for reproducibility of all figures in the paper are given in Appendix A.

2. Stochastic models of tropical vegetation dynamics.

2.1. Microscopic stochastic Staver–Levin model. The Staver–Levin model describes the interaction between savanna trees, forest trees, and grass patches. In this model, grass represents an “open” patch in which new trees can grow, but it also carries fires that limit the expansion of both savanna and forest trees in distinct manners. At the scale of individual trees and patches of grass, the model relies on a few elementary assumptions (see Figure 1):

- (i) Forest trees grow on “open” patches (currently occupied by grass, saplings, or adult savanna trees) at a rate associated with the amount of seeds available on that open patch and therefore related to the density of trees in its vicinity.
- (ii) Similarly, savanna trees grow on grass-occupied patches at a rate that depends on the local density of adult savanna trees.
- (iii) Grass carries fires that kill forest trees and delay the maturation of savanna saplings into adults.¹

¹Later versions of the Staver–Levin model additionally allow adult savanna trees and saplings to be somewhat flammable; we omit this extension here for ease of exposition (cf. [56]), but our methods readily allow the incorporation of this extra feature.

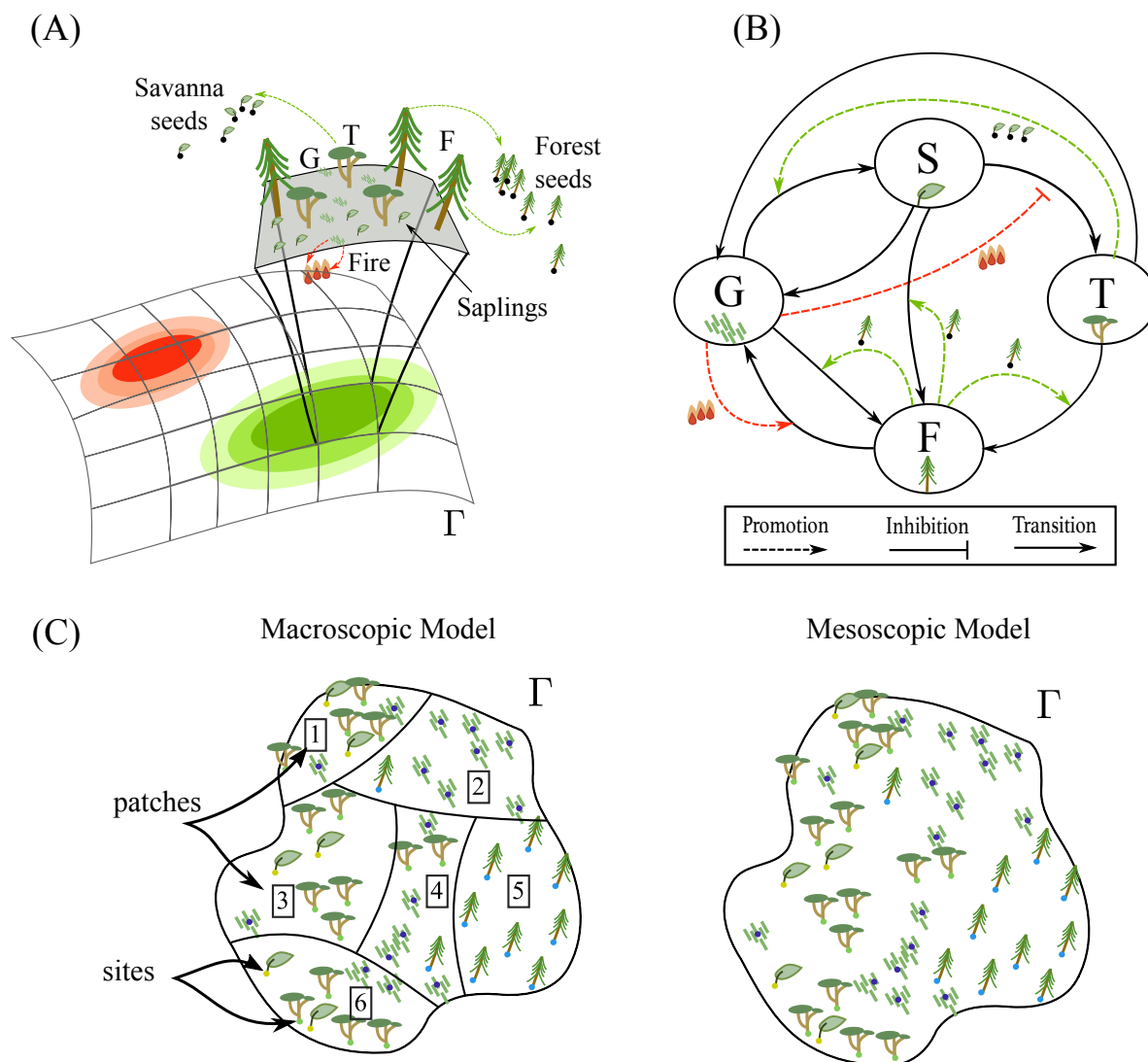


Figure 1. (A) Schematic diagram showing the basic events modeled on a landscape $\Gamma \subset \mathbb{R}^2$: forest trees F (here represented as pine trees for the sake of legibility, although rainforest trees are not coniferous) and savanna trees T disperse seeds carried by the wind, therefore at a limited reach (green ellipses), growing new forest trees or savanna saplings S ; land occupied by no tree is by default covered by grass. Fires, essentially carried by grass, ignite depending on the local density of grass (red ellipses), increase the mortality rate of forest trees, and reduce the rate at which savanna saplings mature into adult trees. (B) Simplified interaction diagram between states in the model (see text). (C) Sample model configurations in the macroscopic model with six patches and eight sites per patch (left) or for the mesoscopic model with randomly located sites (right).

While the original Staver–Levin model does not explicitly consider spatial effects, each of these interactions depends on the density and locations of patches of grass or trees, and extensive remote sensing data highlights the spatial organization at macroscopic scales of savannas and forest lands [47, 48].

To account for spatial interactions, we consider a landscape Γ , which is a Borel set in $\mathbb{R}^2$. On this landscape, multiple *sites*, thought of as small spatial areas of the typical size of a single tree (*microscopic scale*), allow growth of new trees. We suppose that N sites are distributed on the landscape Γ at locations $(r_i)_{i \in \{1, \dots, N\}} \in \Gamma^N$. The locations are assumed to be independent and identically distributed (i.i.d.) according to a probability measure q on Γ . A uniform distribution on Γ is the most natural choice for the r_i 's if we are modeling a homogeneous landscape, but certain heterogeneous environmental features may favor more or less vegetative growth in certain parts of the domain (e.g., soil quality/texture), and this may motivate other choices. If q is a continuous probability measure supported on all of Γ , then we obtain a *mesoscale model* in the limit as $N \rightarrow \infty$, the entire domain is populated with sites, and interactions depend on the precise location of each site (Figure 1(C), right). Some applications emphasizing the interaction between multiple regional covers or populations may motivate choosing a discrete measure for q with a finite number of locations, referred to as patches in this context. In this case, we retain a discrete spatial structure with multiple interacting populations located at each patch, but the number of sites in each patch will still tend to infinity as $N \rightarrow \infty$ (see Figure 1(C), left). This will be referred to as the *macroscale model*.

Each site may switch state at random times with a stochastic intensity that depends on the states of other sites. We denote by $X^i(t)$ the state of site i at time t , and label G the grass state, S the savanna sapling state, T the adult savanna tree state, and F the forest tree state. The transitions of a given site between states are parameterized as follows (see Figure 1(A–B)):

- (i) A savanna sapling grows from a site i currently covered by grass with a rate depending on the total amount of savanna tree seeds available at location r_i . We thus introduce a savanna seed dispersal kernel, denoted $J_S(r, r')$, corresponding to the rate at which a savanna seed travels from r' to r , yielding a transition rate from grass to sapling at site i given by

$$\frac{1}{N} \sum_{j=1}^N J_S(r_i, r_j) \mathbb{1}_{\{X^j(t)=T\}},$$

where $\mathbb{1}_{\{X^j(t)=T\}}$ is the indicator function, equal to 1 when $X^j(t) = T$ and to 0 otherwise. The transition rate is renormalized by the total number of sites to ensure that it remains within fixed bounds as N is varied—this scaling can also be considered as a scaling of time.

- (ii) Similarly, a forest tree grows from a site i currently covered by grass, sapling, or savanna tree (owing to the assumption that forest trees are competitively dominant to both grass and savanna trees alike) with a rate depending on the total amount of seeds available. The number of forest tree seeds available at location r_i depends on a

forest *seed dispersal* kernel $J_F(r, r')$, yielding the transition rate

$$\frac{1}{N} \sum_{j=1}^N J_F(r_i, r_j) \mathbb{1}_{\{X^j(t)=F\}}.$$

- (iii) Forest trees die and are replaced by grassy patches due to either fires fueled by grass, or background nonfire mortality. The rate at which site i of forest type switches to grass thus depends on the flammable cover available in the vicinity of site i and its capacity to transmit fires; this transition rate is given by

$$\phi \left(\frac{1}{N} \sum_{j=1}^N W(r_i, r_j) \mathbb{1}_{\{X^j(t)=G\}} \right).$$

The *forest mortality function* ϕ is a smooth, sharply increasing sigmoidal function of the flammable cover affecting a site at location r_i . The map ϕ incorporates baseline mortality of forest trees $\phi(0) > 0$. More detailed models of fire propagation in savannas have modeled fire propagation using percolation, and previous studies have shown that fire spread in these models has a sharp sigmoidal profile as a function of the flammable cover [45]. Our approach regarding fire propagation thus remains phenomenological and aims to capture the qualitative character of fire spread while retaining analytic tractability.

- (iv) Savanna saplings mature into adult savanna trees. The associated rate of transition is affected by fires that, by top-killing saplings, delay their maturation, and the probability of being affected by a fire depends, again, on the local flammable cover and its capacity to transmit fires. The maturation rate of a sapling into an adult savanna tree is thus given by

$$\omega \left(\frac{1}{N} \sum_{j=1}^N W(r_i, r_j) \mathbb{1}_{\{X^j(t)=G\}} \right),$$

where ω is a sigmoid function (smooth, decreasing, and bounded) quantifying the delayed maturation associated with top-killed saplings. Once more, the sigmoidal form of ω aims to capture the qualitative properties of fire spread without explicitly modeling the underlying fire process.

- (v) Savanna saplings and trees die at constant rates, independent of the state of the system, denoted by μ and ν , respectively.

The transitions for the *finite-size stochastic Stayer–Levin model* outlined above are summarized in Table 1 (along with the corresponding rates for the mean-field limiting process).

The study of the finite-size model defined above is particularly simple. Indeed, these systems are continuous-time finite-state Markov processes (with 4^N possible states for the system) with bounded transition rates. Due to the large dimension of the process, the Kolmogorov equations are impractical to write and solve, but elementary considerations reveal that the system will always tend to the all-grass state. The state where all sites are grass is an absorbing state, since the transition rates from that state to saplings or trees are equal

Table 1

Transition rates in the finite-size stochastic Staver–Levin model (second column) and the corresponding transition rates of the mean-field limiting process (third column). $P_x(t, r) = \mathbb{P}[\bar{X}(t, r) = x]$ denotes the probability that the mean-field process $\bar{X}$ is in state $x \in \{G, S, T, F\}$ at time t and location r .

Transition	Rate in finite-size system	Rate in mean-field limit
$G \rightarrow S$	$\frac{1}{N} \sum_{j=1}^N J_S(r_i, r_j) \mathbb{1}_{\{X^j(t)=T\}}$	$\int_{\Gamma} J_S(r_i, r') P_T(t, r') dq(r')$
$G, S, T \rightarrow F$	$\frac{1}{N} \sum_{j=1}^N J_F(r_i, r_j) \mathbb{1}_{\{X^j(t)=F\}}$	$\int_{\Gamma} J_F(r_i, r') P_F(t, r') dq(r')$
$F \rightarrow G$	$\phi \left(\frac{1}{N} \sum_{j=1}^N W(r_i, r_j) \mathbb{1}_{\{X^j(t)=G\}} \right)$	$\phi \left(\int_{\Gamma} W(r_i, r') P_G(t, r') dq(r') \right)$
$S \rightarrow T$	$\omega \left(\frac{1}{N} \sum_{j=1}^N W(r_i, r_j) \mathbb{1}_{\{X^j(t)=G\}} \right)$	$\omega \left(\int_{\Gamma} W(r_i, r') P_G(t, r') dq(r') \right)$
$S \rightarrow G$	μ	μ
$T \rightarrow G$	ν	ν

to 0 when no forest or savanna trees are present in the system. Moreover, it is easy to see that each state can reach the all-grass absorbing state in a finite number of steps, with a rate bounded away from 0. A classical result of Markov processes thus ensures that, with probability 1, the Markov chain is absorbed in finite time by the all-grass state. Therefore, the stationary dynamics of the finite-size system will always be trivial [14]. This trivial stationary solution sharply contrasts with the complex dynamics of the Staver–Levin model [56]. Thus in section 3.1 we will pay particular attention to the transient dynamics and their duration, two elements of particular interest in ecological modeling that have garnered increasing research interest in recent years [30].

2.2. A general convergence result. The finite-size stochastic Staver–Levin model is one exemplar of a wide class of models arising in ecology and physics in which a large number N of sites (agents, particles, etc.) with fixed spatial locations interact according to the state of their neighbors. Mathematically, we will study a general class of Markovian models where each of the N sites $i \in \{1, \dots, N\}$ can be in one of K states. Each site $i \in \{1, \dots, N\}$ is located at an i.i.d. spatial position r_i drawn from a probability measure q on $\Gamma \in \mathcal{B}(\mathbb{R}^2)$ (where $\mathcal{B}(\cdot)$ denotes the Borel σ -algebra on a given set). The state of each site at time $t \in \mathbb{R}_+$ is denoted $X^i(t)$, and the set of possible states is denoted S^K and has K elements (in the Staver–Levin model, these are the vegetation types $\{G, S, T, F\}$). The initial state of each site is independent of that of all other sites, and each site's initial state has the same space-dependent law satisfying the following regularity condition:

H_{IC} . The initial distribution for site i is given by $\xi^i(r_i)$ where the collection of processes $(\xi^j(r))_{j \in \mathbb{N}}$ are independent random variables indexed by space, with probabilities measurable with respect to space. Rigorously, we assume that their distribution is given

by a Markov kernel $\mu_0 : \Gamma \rightarrow \Omega$ with $(\Omega, \mathcal{F}, \mathbb{P})$ the current probability space.²

Following a classical convention, we consider càdlàg versions of our jump processes (i.e., right-continuous with left limits everywhere), and we denote by $X^i(t^-)$ the left-limit of the state of site i at time t . The transition rate for site i at location r_i and between any two states x and $y \neq x$ depends on the local densities of some other species at a given time t . This rate is thus time dependent in that it varies with the state of the system. Between the last transition before time t and time t , the transition rate below is therefore constant (since no site has changed state) and is assumed to be given by nonlinear functions of a weighted sum of other sites' states:

$$(2.1) \quad R_{x,y}^{i,N}(X(t^-)) = \Phi_{x,y} \left(\frac{1}{N} \sum_{j=1}^N W_{x,y}(r_i, r_j) \mathbb{1}_{\{X^j(t^-)=\psi(x,y)\}} \right),$$

where the kernels $\{W_{x,y} : (x,y) \in S^K \times S^K\}$ weight the nonlocal influence of other sites on the transition rate from state x to state y at a given site. Here, $\Phi_{x,y} : \mathbb{R}_+ \mapsto \mathbb{R}_+$ are smooth (nonlinear) functions, and the maps $\psi(x,y) : S^K \times S^K \mapsto S^K \cup \{\emptyset\}$ are the states that the transition $x \rightarrow y$ depends on. For definiteness, state $\emptyset$ is a state not belonging to S^K , chosen if the transition rate is independent of any other state, e.g., when the transition cannot occur. For simplicity, we assume that transitions depend only on one other species, but it would be straightforward to generalize to transitions depending on multiple species by making the functions $\Phi_{x,y}$ multivariate. In the four-species Stayer–Levin model, the functions $\Phi_{x,y}$ and ψ are given by:

$\Phi_{x,y}(u)$	$x = G$	S	T	F	$\psi(x,y)$	$x = G$	S	T	F
$y = G$	0	ν	μ	$\phi(u)$	$y = G$	$\emptyset$	$\emptyset$	$\emptyset$	G
$y = S$	u	0	0	0	$y = S$	T	$\emptyset$	$\emptyset$	$\emptyset$
$y = T$	0	$\omega(u)$	0	0	$y = T$	$\emptyset$	G	$\emptyset$	$\emptyset$
$y = F$	u	u	u	0	$y = F$	F	F	F	$\emptyset$

Further, we assume the following:

- H1. $\Phi_{x,y} : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is bounded on compacts and Lipschitz continuous for each (x,y) .
- H2. $W_{x,y} : \Gamma \times \Gamma \mapsto \mathbb{R}_+$ is bounded and Borel-measurable for each (x,y) .

Theorem 2.2.1. *Consider the Markov process with rates given by (2.1), with initial conditions satisfying assumption H_{IC} and rates given by (2.1) satisfying assumptions H1 and H2.*

I. Convergence. *For any time $\tau > 0$ and any site $i \in \mathbb{N}$, the process $X^i = \{X^i(t) : t \in [0, \tau]\}$ converges in law, as $N \rightarrow \infty$, to the process $\bar{X}(r_i) = \{\bar{X}(t, r_i) : t \in [0, \tau]\}$ where $\bar{X}(t, r)$ is the unique strong solution to the spatially extended McKean–Vlasov jump process with transitions*

²A Markov kernel (see, e.g., [43, Chap. 3, Def. 1.1]) with source $(X, \mathcal{A})$ and target $(\Omega, \mathcal{F})$, denoted $\kappa : X \rightarrow \Omega$, is a map $\kappa : X \times \Omega \mapsto [0, 1]$ such that

- (i) for every $F \in \mathcal{F}$, the map $x \mapsto \kappa(x, F)$ is $\mathcal{A}$ -measurable;
- (ii) for every $x \in A$, the map $\kappa(x, \cdot)$ is a probability measure.

In other words, κ associates to every point x a probability measure such that for every measurable set of the probability space, the probability $x \mapsto \kappa(B, x)$ is measurable.

occurring at independent, exponentially distributed times with rates given by

$$(2.2) \quad x \rightarrow y \quad \Phi_{x,y} \left(\int_{\Gamma} W_{x,y}(r, r') \mathbb{P}[\bar{X}(t^-, r') = \psi(x, y)] dq(r') \right)$$

for each pair of states $(x, y) \in S^K \times S^K$ and initial condition with law μ_0 .

II. Characterization of the limit. Define $P_x(t, r) = \mathbb{P}[\bar{X}(t, r) = x]$ for each $x \in S^K$. The probability distribution of the process $\bar{X}$ satisfies the system of GKEs,

$$(2.3) \quad \partial_t P_x(r) = \sum_{y \neq x} \Phi_{y,x} \left(\int_{\Gamma} W_{y,x}(r, r') P_{\psi(y,x)}(r') dq(r') \right) P_y(r) \\ - \sum_{y \neq x} \Phi_{x,y} \left(\int_{\Gamma} W_{x,y}(r, r') P_{\psi(x,y)}(r') dq(r') \right) P_x(r), \quad x \in S^K,$$

where $P_y(r)$ is shorthand for $P_y(t^-, r)$ for each $y \in S^K$ and each $(t, r) \in \mathbb{R}_+ \times \Gamma$.

III. Propagation of chaos. Moreover, any fixed finite subset of sites $(X^{i_1}, \dots, X^{i_p})$ converges towards independent variables; i.e., for any collection of p states $(x_1, \dots, x_p)$,

$$\lim_{N \rightarrow \infty} \mathbb{P}[X^{i_1}(t) = x_1, \dots, X^{i_p}(t) = x_p] = \mathbb{P}[\bar{X}(t, r_{i_1}) = x_1] \times \dots \times \mathbb{P}[\bar{X}(t, r_{i_p}) = x_p]$$

for each $t \in [0, \tau]$.

Theorem 2.2.1 demonstrates that the Markov process given by (2.1) is well approximated in the large N limit by the stochastic jump process $\bar{X}$, which we refer to as the mean-field limit. $\bar{X}$ is not a Markovian jump process: the rate of transition depends explicitly on the law of the process $\bar{X}$ and not only on its current state. These types of processes are referred to as McKean–Vlasov processes, and it is generally hard to characterize properties of their solutions [58]. However, in this case, Theorem 2.2.1 provides a simple characterization of the solution in the form of a system of either

- IDEs when q is a continuous distribution (mesoscopic model), or
- nonlinear ODEs when q is a finite combination of Dirac masses (macroscale model).

We refer to the system of equations given by (2.3) as the *generalized Kolmogorov equations* (GKEs) for $\bar{X}$ because, similar to classical Kolmogorov equations for Markovian processes, they describe the time evolution of the probability density of the continuous-time finite-state jump process $\bar{X}$. However, they are not standard Kolmogorov equations, since $\bar{X}$ is non-Markovian—this is due to the dependence of $\bar{X}$'s transition rates on the probability density of the process itself. This dependence makes the GKEs nonlinear, contrasting with classical Kolmogorov equations. The GKEs are also more analytically tractable than the finite-size Markov process (2.1)—or even the linear Kolmogorov equations associated to the finite-size system, since these equations are very high dimensional (associated with a K^N system of linear differential equations). Furthermore, the GKEs naturally satisfy a conservation law that was included in the original Staver–Levin model [49], i.e.,

$$(2.4) \quad \sum_{x \in S^K} P_x(t, r) = 1 \quad \text{for each } r \in \Gamma \quad \text{and } t \in [0, \tau].$$

The proof of convergence in Theorem 2.2.1 uses a coupling method consisting of two steps. First, we prove strong existence and uniqueness of solutions to the associated mean-field equation given by (2.2). Second, we construct a particular solution $\bar{X}^i$ of the limiting equation having the same initial condition and coupling the jump times with those of the finite system, to which the process $\{X^i(t), t \in [0, \tau]\}$ converges almost surely. The rate of convergence is also quantified through this coupling argument, and it is shown that the distance between the finite system and their limit decays as $1/\sqrt{N}$.

The coupling methods we employ are classical and have traditionally been applied to high dimensional Markov processes, often with multiple classes, but typically without spatial structure [1, 16, 15, 24, 29]. It is only recently that we extended these methods to spatial systems with uncountable state spaces (i.e., interacting particle systems driven by independent Brownian motions) [53, 54]. Some recent works have addressed the question of mean-field convergence with spatial structure but only for nearest neighbor interactions [51]. The Stayer–Levin model specifically motivates the extension of these methods to finite-state continuous-time Markov processes with nonlocal interactions, due to the nonlocal and long-range nature of seed dispersal [9, 52].

A direct application of Theorem 2.2.1 shows that the stochastic Stayer–Levin model with the rates given by Table 1 converges towards a non-Markovian spatially extended McKean–Vlasov jump process with transition rates given by

$$(2.5) \quad \begin{cases} G \rightarrow S, & \int_{\Gamma} J_S(r, r') P_T(t, r') dq(r'), \\ G, S, T \rightarrow F, & \int_{\Gamma} J_F(r, r') P_F(t, r') dq(r'), \\ F \rightarrow G, & \phi \left(\int_{\Gamma} W(r, r') P_G(t, r') dq(r') \right), \\ S \rightarrow T, & \omega \left(\int_{\Gamma} W(r, r') P_G(t, r') dq(r') \right), \\ S \rightarrow G, & \mu, \\ T \rightarrow G, & \nu. \end{cases}$$

In this case, the GKEs of the mean-field limiting process with transition rates given by (2.5) are

$$(2.6) \quad \begin{cases} \partial_t P_G(r) &= \mu P_S(r) + \nu P_T(r) + \phi \left(\int_{\Gamma} W(r, r') P_G(r') dq(r') \right) P_F(r) \\ &\quad - P_G(r) \int_{\Gamma} J_S(r, r') P_T(r') dq(r') - P_G(r) \int_{\Gamma} J_F(r, r') P_F(r') dq(r'), \\ \partial_t P_S(r) &= P_G(r) \int_{\Gamma} J_S(r, r') P_T(r') dq(r') - P_S(r) \int_{\Gamma} J_F(r, r') P_F(r') dq(r') \\ &\quad - P_S(r) \omega \left(\int_{\Gamma} W(r, r') P_G(r') dq(r') \right) - \mu P_S(r), \\ \partial_t P_T(r) &= -\nu P_S(r) - P_T(r) \int_{\Gamma} J_F(r, r') P_F(r') dq(r') \\ &\quad + P_S(r) \omega \left(\int_{\Gamma} W(r, r') P_G(r') dq(r') \right), \\ \partial_t P_F(r) &= (1 - P_F(r)) \int_{\Gamma} J_F(r, r') P_F(r') dq(r') \\ &\quad - \phi \left(\int_{\Gamma} W(r, r') P_G(r') dq(r') \right) P_F(r). \end{cases}$$

Theorem 2.2.1 thus provides a natural candidate for a deterministic spatially extended Staver–Levin model, distinct from classical extensions generally done for ecological models and relying on diffusion operators [39]. Our framework also allows consideration of situations different from those covered by Durrett and collaborators [18, 20, 21], in particular regarding arbitrary site distributions and arbitrary smooth kernels and spatial dependence, although our approach is limited to smooth dependence on the states (Lipschitz assumptions H1 and H2). Furthermore, owing to their relationship with a fine microscopic model, the IDEs (2.6) are ecologically relevant: the spatial integral terms can be readily interpreted as nonlocal effects of seed dispersal and fire propagation.

Example 2.2.2 (macroscale vegetation model and the Staver–Levin model as a metapopulation or network model). Macroscale models correspond to cases where the landscape is described as composed of a finite number $M \in \mathbb{N}$ of patches $j \in \{1, \dots, M\}$. This case is treated under the consideration that each patch is labeled through a location variable in the space $\mathcal{A}_M := \{1, \dots, M\}$. Now suppose that $q = (q_j)_{j=1, \dots, M}$ is a probability measure on $\mathcal{A}_M$, and denote by P_x^j the probability that the population in patch j is in state $x \in \{G, S, T, F\}$. By Theorem 2.2.1, these probabilities satisfy the following GKEs in the limit as $N \rightarrow \infty$:

$$(2.7) \quad \begin{cases} \frac{dP_G^j}{dt} = \mu P_S^j + \nu P_T^j + \phi \left(\frac{1}{M} \sum_{k=1}^M q_k W(j, k) P_G^k \right) P_F^j \\ \quad - \frac{P_G^j}{M} \sum_{k=1}^M q_k J_S(j, k) P_T^k - \frac{P_G^j}{M} \sum_{k=1}^M q_k J_F(j, k) P_F^k, \\ \frac{dP_S^j}{dt} = \frac{P_G^j}{M} \sum_{k=1}^M q_k J_S(j, k) P_T^k - \frac{P_S(r)}{M} \sum_{k=1}^M q_k J_F(j, k) P_F^k \\ \quad - P_S^j \omega \left(\frac{1}{M} \sum_{k=1}^M q_k W(j, k) P_G^k \right) - \mu P_S^j, \\ \frac{dP_T^j}{dt} = -\nu P_S^j - \frac{P_T^j}{M} \sum_{k=1}^M q_k J_F(j, k) P_F^k + P_S^j \omega \left(\frac{1}{M} \sum_{k=1}^M q_k W(j, k) P_G^k \right), \\ \frac{dP_F^j}{dt} = \frac{1}{M} (1 - P_F^j) \sum_{j=1}^M q_k J_F(j, k) P_F^k - \phi \left(\frac{1}{M} \sum_{k=1}^M q_k W(j, k) P_G^k \right) P_F^j \end{cases}$$

for each $j \in \{1, \dots, M\}$. If $M = 1$ and

$$W(1, 1) = 1, \quad J_S(1, 1) = \beta, \quad J_F(1, 1) = \alpha,$$

the GKEs are exactly the classical Staver–Levin ODE model [49],

$$(2.8) \quad \begin{cases} \dot{G} = \mu S + \nu T + \phi(G) F - \beta G T - \alpha G F, \\ \dot{S} = \beta G T - \alpha G F - S \omega(G) - \mu S, \\ \dot{T} = -\nu S - \alpha T F + \omega(G) S, \\ \dot{F} = \alpha (1 - F) F - \phi(G) F, \end{cases}$$

where G , S , T , and F denote the proportions of landscape covered by grass, saplings, savanna trees, or forest trees in the original Staver–Levin model, while in our context they represent

the probability for a given site to be in one of these states. In other words, the condition imposed in the original Stayer–Levin model ensuring that the variables are proportions, i.e.,

$$(2.9) \quad G(t) + S(t) + T(t) + F(t) = 1 \quad \text{for all } t \geq 0,$$

corresponds exactly to the probabilistic constraint (2.4) obeyed by solutions to the GKEs. Because of the propagation of chaos property (Theorem 2.2.1 III), finite subsets of sites become independent in the large N limit, and therefore the empirical proportion of landscape covered by a given species on this subset is well approximated by the probability for any given site to be in that state.

Example 2.2.3 (homogeneous isotropic grass-forest model). We consider an example of competition between grass and forest (without savanna trees) on a landscape Γ , a Borel set in $\mathbb{R}^2$. We assume that the landscape is homogeneous (i.e., q is the uniform measure on Γ). Furthermore, as is typical in applications, we assume that W and J_F are radial, i.e., $f(r, r') = \hat{f}(|r - r'|)$ for $f = J_F$ or W , and where $|\cdot|$ denotes the Euclidean norm on Γ . Theorem 2.2.1 implies that the GKEs associated with the mean-field limit in this case are

$$(2.10) \quad \begin{cases} \partial_t P_G(r) &= \phi \left((\hat{W} * P_G)(r) \right) P_F(r) - P_G(r) (\hat{J}_F * P_F)(r), \\ \partial_t P_F(r) &= (1 - P_F(r)) (\hat{J}_F * P_F)(r) - \phi \left((\hat{W} * P_G)(r) \right) P_F(r) \end{cases}$$

for each $r \in \Gamma$, with $(f * g)(r) = \int_{\Gamma} f(|r - r'|)g(r') \frac{dr'}{\text{Leb}(\Gamma)}$ with $\text{Leb}(\Gamma)$ denoting the Lebesgue measure of Γ . The system of equations (2.10) is also subject to the constraint

$$P_G(t, r) + P_F(t, r) = 1 \quad \text{for each } r \in \Gamma \quad \text{and } t \in [0, \tau].$$

3. Spatially extended Stayer–Levin model: Analysis of solutions. Our finite-size Stayer–Levin models are Markov processes with an absorbing (or trapping) state that can be reached from any other state. Therefore, the unique stationary solution for these processes is the all-grass state, and, with probability 1, this state is reached in finite time. However, this absorption can arise after long transients; this transient behavior is particularly important in ecology and other applications, where time scales are typically long [30]. In particular, this finite-time absorption is in sharp contrast with the remarkable complexity of the original Stayer–Levin model, where the all-grass state may be unstable and other equilibria or periodic orbits emerge [56]. Thus it is of evident interest to study the consistency between finite-size tree extinction and nonextinction in the mean-field limit for both the macroscale and mesoscale frameworks. All details related to numerics and coding for this section can be found in Appendix A.

3.1. Macroscopic Markov model and generalized Stayer–Levin model. We first investigate absorption properties and the consistency between the quasi-stationary distributions (behavior of the finite-size system prior to extinction) and the GKEs of the mean-field limiting process. We then demonstrate long-run transient periodic solutions to the particle system in a parameter regime for which the limiting process has a periodic law.

3.1.1. Mean-field behavior in a single patch. We analyze in detail the absorption property of the macroscale model introduced in Example 2.2.2 with a single patch in the forest-grass subsystem. In the notation of Example 2.2.2, suppose

$$W(1, 1) = 1, \quad J_F(1, 1) = \bar{J} > 0.$$

Since we restrict our attention to the forest-grass subsystem, the finite-size system X is a classical two-state Markov process in dimension N with (for each site $i \in \{1, \dots, N\}$) independent exponentially distributed transitions with rates given by

$$(3.1) \quad \begin{cases} G \rightarrow F, & \bar{J} \hat{P}_F^N(t) & \text{(forest tree seed dispersal),} \\ F \rightarrow G, & \phi(1 - \hat{P}_F^N(t)) & \text{(burning of forest trees),} \end{cases}$$

where $\hat{P}_F^N = N^{-1} \sum_{j=1}^N \mathbb{1}_{\{X^j(t)=F\}}$ is the fraction of forest sites (or empirical probability of the forest state across all sites) and $\bar{J}$ is the birth rate of forest trees. The mean-field limit $\bar{X}$ is a two-state non-Markovian process with independent exponentially distributed transitions that occur with rates

$$(3.2) \quad \begin{cases} G \rightarrow F, & \bar{J} P_F(t), \\ F \rightarrow G, & \phi(1 - P_F(t)), \end{cases}$$

where $P_F(t)$ and $P_G(t)$ are the probabilities that $\bar{X}$ is in the forest state or grass state, respectively. These probabilities can be computed by solving the GKEs of the mean-field limit,

$$(3.3) \quad \frac{d}{dt} P_G(t) = (1 - P_G(t)) \phi(P_G(t)) - \bar{J} P_G(t) (1 - P_G(t))$$

with $P_G(t) = 1 - P_F(t)$. If $\phi(0) > 0$ and $\bar{J} > 0$, the all-grass state is the unique stationary distribution for the finite-size particle system, and, from any initial distribution, absorption occurs in finite time with probability 1. Consistent with this, $P_G = 1$ is always a fixed point of the GKEs for the corresponding mean-field limit. Moreover, this fixed point is attractive for forest tree birth rate small enough but loses stability when the tree birth rate exceeds tree death rate in the all-grass state (i.e., $\bar{J} < \phi(1)$). Therefore, for larger birthrates, the mean-field limit will not converge towards the all-grass state, the unique attractor of the finite system.

The bifurcation diagram for the GKEs of the mean-field limit process with rates given by (3.2) is computed as a function of the forest tree birth rate $\bar{J}$ (Figure 2, red); it features two saddle-nodes and a transcritical bifurcation that split the parameter space into four main regimes:

- For forest birth rate small enough, the only stable equilibrium is the all-grass state.
- A saddle-node bifurcation (arising at $\bar{J} \approx 0.55$ for our choice of parameters) leads to the emergence of a stable forest-dominated equilibrium as well as an unstable fixed point. The all-grass equilibrium conserves its stability, but, depending on the initial grass cover relative to the unstable equilibrium, the system converges to either the all-grass solution or the stable forest-dominated equilibrium.

- As the forest birthrate is further increased, the all-grass state loses stability (at $\bar{J} = \phi(1)$) in favor of a grass-dominated equilibrium through a transcritical bifurcation. In that case, whatever the initial condition, the limit system never reaches the all-grass equilibrium. Moreover, as $\bar{J}$ is increased, the grass cover of the grass-dominated equilibrium decreases progressively as the forest tree birthrate increases, before disappearing through a second saddle-node bifurcation.
- Beyond this second saddle-node bifurcation, the only stable equilibrium persisting is the forest-dominated equilibrium.

The blue stars in Figure 2 show the state of the finite-size system at time $t = 100$ from simulations of the one-patch two-state Markov chain with $N = 3000$ sites for various values of the forest birthrate $\bar{J}$. We simulated the finite-size system for each value of $\bar{J}$ for various initial conditions with various initial grass cover proportions (20 initial conditions when the GKEs are monostable, and 90 initial conditions regularly sampled between 0.1 and 0.9 in the bistable case). There is remarkable agreement between the simulations of the Markov process and the bifurcation diagram in Figure 2. Even after a reasonably long time ($t = 100$), the finite-size system remained away from the all-grass state for many initial conditions when other stable attractors existed in the GKEs of the mean-field limit. Moreover, matching the initial grass cover with the final state, we observed that the “basin of attraction” of each quasi-attractor for the transient states of the finite-size Markov process system shows excellent agreement with the unstable fixed point from the GKEs. Past the transcritical bifurcation we continued to observe absorption to the all-grass state, although this fixed point is unstable in the GKEs for this parameter range.

3.1.2. Quasi-stationary distribution and absorption rate. The apparently paradoxical persistence of transient behavior and striking agreement between the mean-field and finite-size systems in Figure 2 can be better understood in terms of the quasi-stationary distribution (QSD) of the finite-size system. Fix $N \in \mathbb{N}$, and consider the Markov process $\tilde{X}$ which tracks the proportion of grass sites in the finite-size system so that the state space of $\tilde{X}$ is $S = \{1, (N-1)/N, \dots, 1/N, 0\}$. The QSD is the stationary distribution of the system conditional on not being absorbed by the all-grass state or, equivalently, it is the stationary distribution of the new process $\tilde{X}^*$ on the restricted state space $S^* = \{(N-1)/N, \dots, 1/N, 0\}$.

Definition 3.1.1. A distribution $x \in \mathbb{R}^N$ is a QSD for the process $\tilde{X}$ if for each $t \geq 0$,

$$\mathbb{P}_x \left[\tilde{X}(t) = j \mid T > t \right] = x_j, \quad j \in S^*,$$

where $T = \inf\{t \geq 0 : \tilde{X}(t) = 1\}$, and $\mathbb{P}_{x_0}$ denotes the probability measure associated with the process $\tilde{X}$ starting from the initial distribution x_0 .

The generator matrix of the process $\tilde{X}$ has the form

$$\begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{a} & Q \end{pmatrix},$$

where $\mathbf{0}^T = \{0, \dots, 0\} \in \mathbb{R}^N$,

$$\mathbf{a} = \left\{ \frac{1}{N} \phi \left(\frac{N-1}{N} \right), 0, \dots, 0 \right\}^T \in \mathbb{R}^N,$$

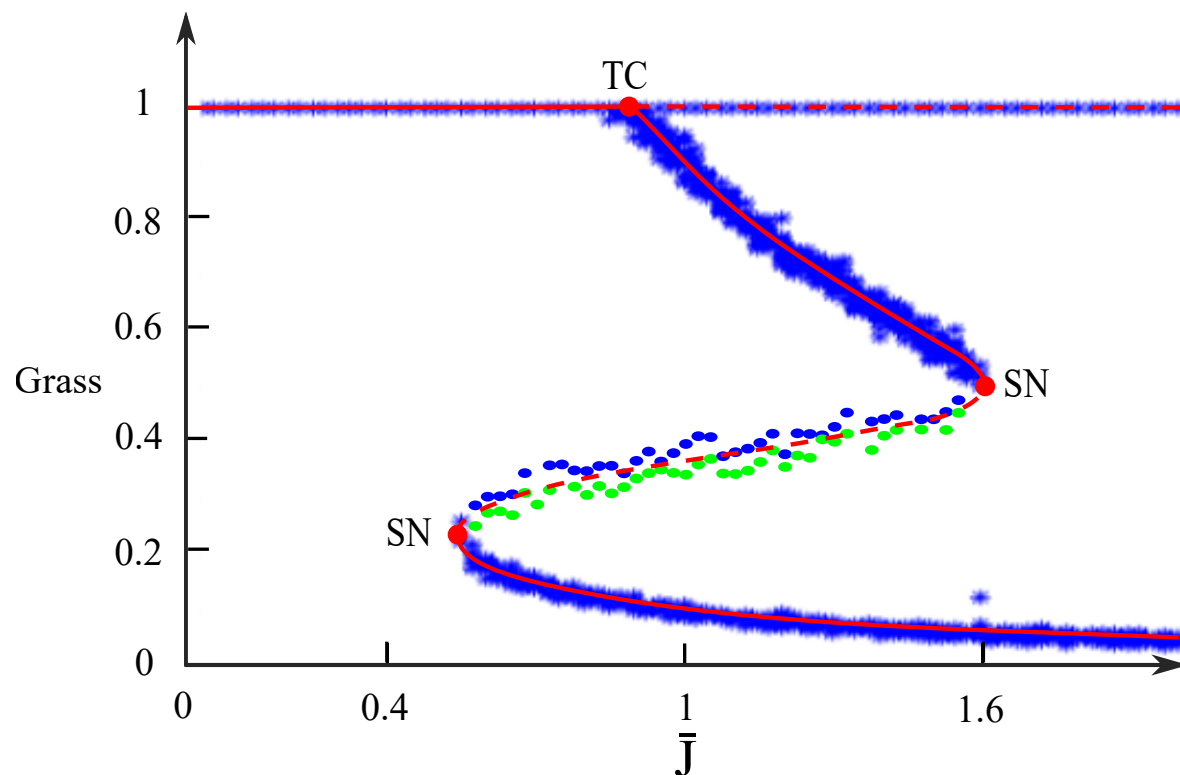


Figure 2. *Blue starred markers:* Grass proportion in the finite-size Staver–Levin model for time $t = 100$ with $N = 3000$ and $M = 1$ (macroscale model). End states from 20 different initial conditions are plotted for each value of the forest recruitment rate $\bar{J}$, and the attraction basin was computed using 90 different initial conditions for more precision. *Red curves:* The bifurcation diagram of the GKEs for the corresponding mean-field limit is overlaid with solid red lines indicating stable equilibria and dashed red lines indicating unstable equilibria. *Blue/green dots:* Blue dots mark the lowest initial condition for the finite-size system that resulted in a final state on the upper equilibrium branch, and green dots mark the largest average initial condition which gave an end state on the lower equilibrium branch. *Red dots:* These markers indicate the bifurcation points in the GKEs of the mean-field limit (SN: saddle-node bifurcation, TC: transcritical bifurcation).

and Q is the $N \times N$ matrix given by

$$\begin{pmatrix} -\frac{1}{N}\phi\left(\frac{N-1}{N}\right) - \frac{\bar{J}}{N}\frac{N-1}{N} & \frac{\bar{J}}{N}\frac{N-1}{N} & 0 & \cdots & 0 \\ \frac{2}{N}\phi\left(\frac{N-2}{N}\right) & -\frac{2}{N}\left(\phi\left(\frac{N-2}{N}\right) + \bar{J}\frac{N-2}{N}\right) & \frac{2\bar{J}}{N}\frac{N-2}{N} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \cdots & \cdots & \frac{N-1}{N}\phi\left(\frac{1}{N}\right) & -\frac{N-1}{N}\left(\phi\left(\frac{1}{N}\right) + \frac{\bar{J}}{N}\right) & \frac{(N-1)\bar{J}}{N}\frac{1}{N} \\ 0 & 0 & \cdots & 0 & \phi(0) & -\phi(0) \end{pmatrix}.$$

Since $\phi(0) > 0$ and $\bar{J} > 0$, $\tilde{X}^*$ is recurrent on S^* , and hence the QSD exists and is unique [14, 41]. The QSD x associated with the process $\tilde{X}$ is given by

$$xQ = -\rho x, \quad x\mathbf{1}^T = 1,$$

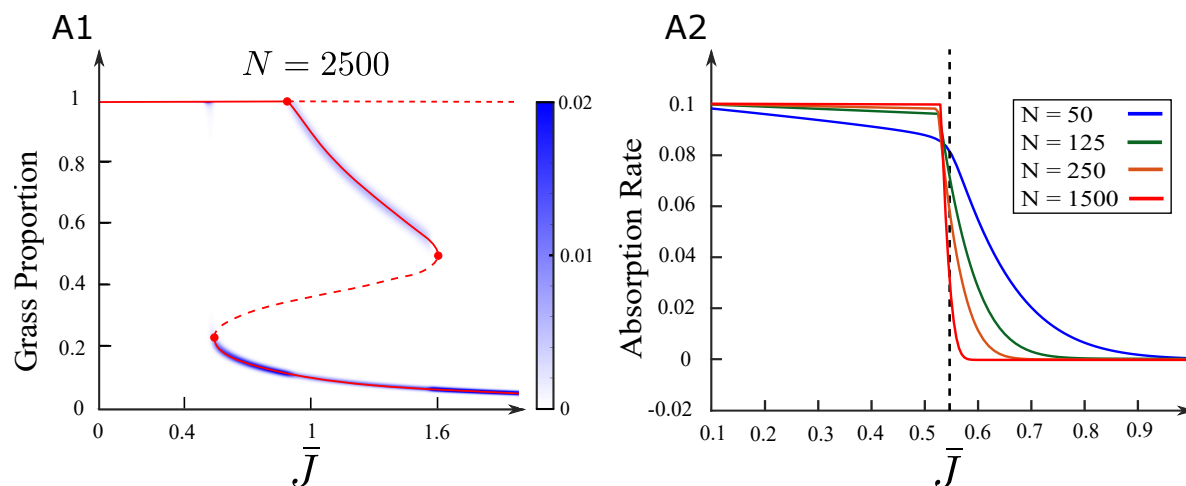


Figure 3. A1: Quasi-stationary distribution of the interacting particle system for $N = 2500$ for each value of $\bar{J}$. A2: Absorption rate to the all-grass state as a function of $\bar{J}$.

where ρ is the dominant eigenvalue of Q (i.e., eigenvalue with largest real part). In other words, the QSD is the normalized eigenvector associated with the principal eigenvalue of Q .

The QSD for the two-state finite-size Stayer–Levin system given by (3.1) is plotted in panel A1 of Figure 3 for each value of $\bar{J}$. The mass of the QSD is concentrated on the stable states of the GKEs for the corresponding mean-field limit—partly explaining the persistent transients observed in Figure 2. A more complete explanation of this persistence is obtained by calculating the absorption rate for the Markov jump process $\tilde{X}$ —this is the speed at which we expect the particle system to approach the all-grass state from an arbitrary initial distribution. More precisely, the dominant eigenvalue of the substochastic restricted transition matrix Q , denoted by ρ , gives the speed of approach to the all-grass state in the sense that for any initial distribution x_0 ,

$$\lim_{t \rightarrow \infty} \mathbb{P}_{x_0} [\tilde{X}(t+s) \neq 1 \mid T > t] = e^{-\rho s}, \quad s \geq 0.$$

Panel A2 of Figure 3 shows the absorption rate ρ as a function of $\bar{J}$ for multiple values of the system size N . The absorption rate shows a decreasing profile, indicating (as expected) a decay of the rate of absorption as the forest tree birth rate increases. More strikingly, the rates decay with a sharpening profile as N increases, with a consistent switch from a rate around 0.1 to an almost zero rate at $\bar{J} \approx 0.55$ —coinciding exactly with the appearance of a saddle-node bifurcation in the GKEs of the mean-field system (cf. Figure 2). Therefore, consistent with Figure 2, we expect to observe persistent transient behavior matching the QSD for values of $\bar{J}$ past the first saddle-node bifurcation.

3.1.3. Stochastic oscillations: Solutions with time periodic law. The original Stayer–Levin model given by the system of nonlinear ODEs (2.8) has regions of parameters where the only stable attractor is a periodic orbit. In this regime, a simple consequence of Theorem 2.2.1 is that the mean-field limiting process will exhibit stable oscillations. This is a remarkable

property and has previously been observed in continuous-time Markov processes described by Brownian driven stochastic differential equations (SDEs) [46, 55]. To the best of our knowledge, the mean-field process defined by (2.5) is the only example identified to date of a McKean–Vlasov jump process which has a periodic law. Mathematical methods for studying nonstationary solutions are still to be developed for jump processes, but in our case the existence of a periodic law is based on the derivation of the GKEs and their bifurcation analysis, avoiding the need for probabilistic arguments.

Panels A1 and A2 of Figure 4 show the periodic solutions generated by the four-species macroscale model introduced in Example 2.2.2 in a periodic regime with $M = 1$. We observe, consistent with Theorem 2.2.1, that (for N sufficiently large) trajectories of the finite-size Staver–Levin system remain close to the trajectories of the GKEs of the corresponding mean-field limit, a result analogous to the effect observed in [19, Figure 4]. Periodic orbits in the original Staver–Levin model were shown to grow and disappear at a heteroclinic orbit when the forest tree birth rate increases [56]. We thus explored the dynamics of the finite-size Staver–Levin model as a function of the forest tree birth rate $\bar{J}$. The heteroclinic orbit in the original Staver–Levin model connects three fixed points: the all-grass equilibrium, a savanna equilibrium, and a mixed-saddle equilibrium where all species are present. We simulated trajectories of the finite-size system for various values of $\bar{J}$ (see Figure 4(C2)) and found that while transient trajectories show dynamics very similar to the GKEs of the mean-field limit, oscillations became transient for parameters too close to the heteroclinic cycle; the system is rapidly absorbed by the savanna subsystem (i.e., absence of forest trees) and reaches a fixed point on this subsystem. Rigorously, this fixed point is unstable for the GKEs because of an invasion of forest trees, and trajectories of the (deterministic) GKEs visit regions very close to that fixed point. The finite-size system, following closely these trajectories, thus reaches states with very low numbers of forest trees where extinction of forest becomes very likely. Near heteroclinic cycles the finite-size system is thus vulnerable to fluctuations in the vicinity of absorbing states. Interestingly, this effect of absorption near the heteroclinic cycle shows dynamics significantly different from simulations of the GKEs with Brownian noise (which lead to stochastic resonance phenomena) [56]. From an ecological viewpoint, this reveals an interesting fragility of the ecosystems when trajectories approach absorbing subsystems.

3.2. Mesoscopic model and spatially extended Staver–Levin model. In our mesoscopic framework, we derived a mean-field spatially extended jump process characterizing the dynamics of sites distributed in space. The distribution of this process is described by IDEs which characterize the spatio-temporal dynamics of vegetation at this scale (the GKEs). The rigorous derivation of this model from first principles provides us with a new model that incorporates heterogeneity in the sites' density via the choice of the initial site distribution q . The study of this system can yield valuable information on the distribution of vegetation in space, but we defer a full investigation of this model's dynamics to future work. Here, we concentrate solely on convergence properties and dynamical consistency between the mesoscopic finite-size Staver–Levin model and the GKEs of the corresponding mean-field limit.

We focus on the simplest subsystem possible, the forest-grass subsystem, in order to illustrate our convergence results. In this subsystem, the dynamics are fully described by the

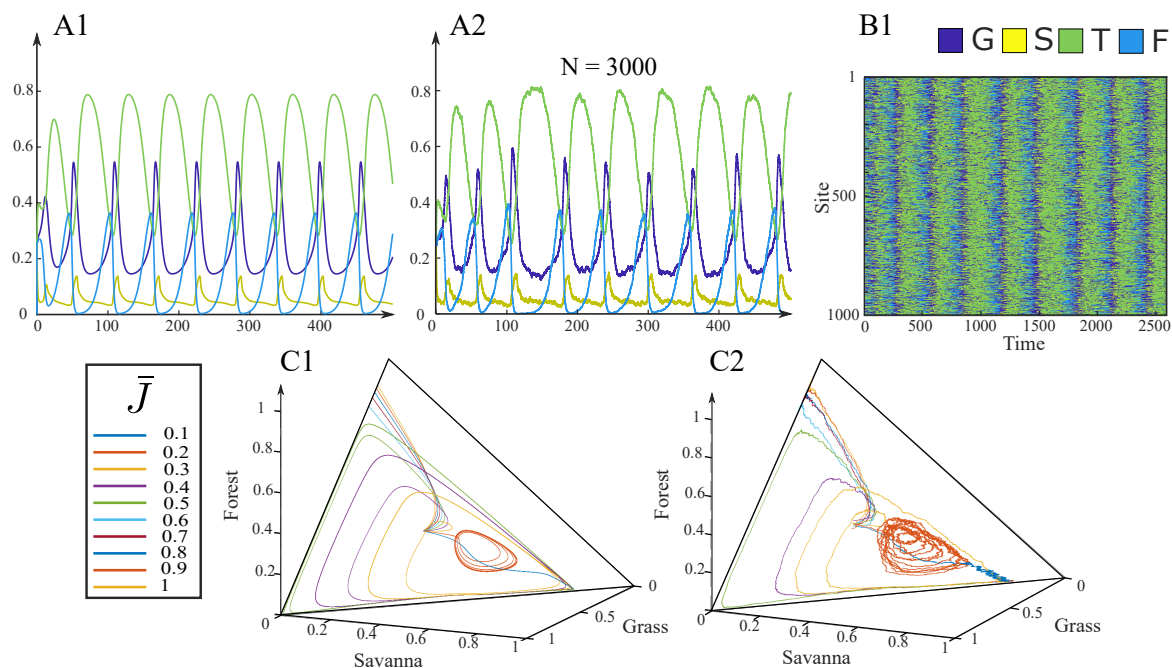


Figure 4. A1/A2: Comparison of the single patch macroscopic particle system model for $N = 3000$ (A2) and the solution of the corresponding Kolmogorov equations in a periodic parameter regime. B1: Simulation of the macroscopic model for a large number of sites in a periodic regime. C1/C2: Comparison of trajectories of the particle system and Kolmogorov equation solutions in the state space for various values of α .

fraction of forest trees at a given location $r \in \Gamma$ and time $t \in \mathbb{R}^+$, i.e., the quantity

$$P_F(t, r) := \mathbb{P} [\bar{X}(t, r) = F].$$

The GKEs of the mean-field process are thus given by the following nonlocal IDE:

$$(3.4) \quad \frac{\partial}{\partial t} P_F(t, r) = P_G(t, r) \int_{\Gamma} J(r', r) P_F(t, r') dq(r') - P_F(t, r) \phi \left(\int_{\Gamma} W(r', r) P_G(t, r') dq(r') \right), \quad (t, r) \in \mathbb{R}_+ \times \Gamma,$$

with $P_G(r, t) = 1 - P_F(r, t)$. We study two ecologically relevant behaviors arising in these systems, waves of invasion and emergence of fronts in heterogeneous landscapes, as well as the consistency between solutions of the finite-size system and the IDE (3.4).

3.2.1. Waves of invasion. Consider a homogeneous one-dimensional landscape, assumed for simplicity to be a ring (i.e., $\Gamma = \mathcal{S}_L$, the one-dimensional torus of length L , represented by the interval $[0, L]$ with the boundaries identified). Assuming a homogeneous landscape amounts to considering a uniform site density q on Γ within our framework. Both seed

dispersal and fire propagation kernels are Gaussian functions of the form

$$W(r, r') = \frac{J(r, r')}{\bar{J}} = \frac{C(\sigma)}{\sigma\sqrt{2\pi}} e^{-\frac{d_\Gamma(r, r')^2}{2\sigma^2}}, \quad (r, r') \in \Gamma^2,$$

where $d_\Gamma(r, r')$ denotes the distance between the points r and r' on the ring Γ . The nonstandard normalization factor $C(\sigma) = L(2\Phi(L/2\sigma) - 1)^{-1}$ is due to the fact that our Gaussian kernels are compactly supported and integrated against the uniform measure on $[0, L]$.³ In the present setup, spatially homogeneous solutions of the IDE (3.4) solve the original Staver–Levin ODEs (2.8) due to the translation invariance on the ring. Therefore, spatially homogeneous stationary solutions are given by the bifurcation diagram of the ODE in Figure 2, but spatial interactions may alter the stability of these steady states. Grassland is the unique spatially homogeneous equilibrium when the forest tree birth rate $\bar{J}$ is low enough. Forest is the unique spatially homogeneous steady state for sufficiently large $\bar{J}$, while multiple homogeneous steady states coexist for intermediate values of $\bar{J}$. We studied the competition between forest and grass in the coexistence regime by choosing an initial state with a region of forest trees flanked by grass on either side. In Figure 5 (panels A1, A2, A4, and A5) we observe that, depending on the forest tree birth rate $\bar{J}$, grass may invade the patch of forest trees or trees may invade the grassy regions (in both the finite-size system and the GKEs). There is excellent agreement between solutions of the GKEs of the mean-field limit (i.e., the IDEs given by (3.4)) and those of the finite-size Markov process.

3.2.2. Forest-grass fronts in heterogeneous landscapes. One advantage of our framework is the ability to derive mesoscopic models which incorporate environmental heterogeneity by choosing a nonuniform initial site distribution. For instance, ecologically, soils may have substantial effects on tree establishment potential, which can be reflected in our model via lower site density regions. Reduced site density induces two opposite effects: a lessened ability to carry fires and a lower opportunity for trees to grow. The results of simulations of the finite-size model, as well as the GKEs of the corresponding mean-field model, with a variable density of sites in the forest-grass subsystem are presented in Figure 5, panels B1–B3. For these simulations, the site density was chosen to be a trapezoid on $\Gamma = [0, L]$, i.e.,

$$dq(x) = (a + bx) \mathbb{1}_{[0, L]}(x) dx, \quad a, b > 0.$$

This choice ensures an increasing density of sites along the interval Γ : regions near $x = 0$ have densities lower than those of regions near $x = L$. Space-time plots of the solutions to the finite-size model and the GKEs (panels B1 and B2) illustrate that lower site density favors forest, while higher densities favor grassland. At intermediate site densities, a sharp front forms between forest and grassland as neither species is able to invade the other, akin to a Maxwell point. This type of solution is also referred to as a “front-pinning phenomenon” as the sharp front forms because the wave speed of the wave of invasion, which would typically annihilate the less competitive species (in the corresponding homogeneous domain problem), approaches zero at the Maxwell point [57, 59]. Panel B3 compares the bifurcation diagram of the appropriate GKEs without spatial interaction ($\sigma \rightarrow 0^+$ in the kernels) with the final

³ Φ denotes the cumulative distribution function of a standard normal random variable here.

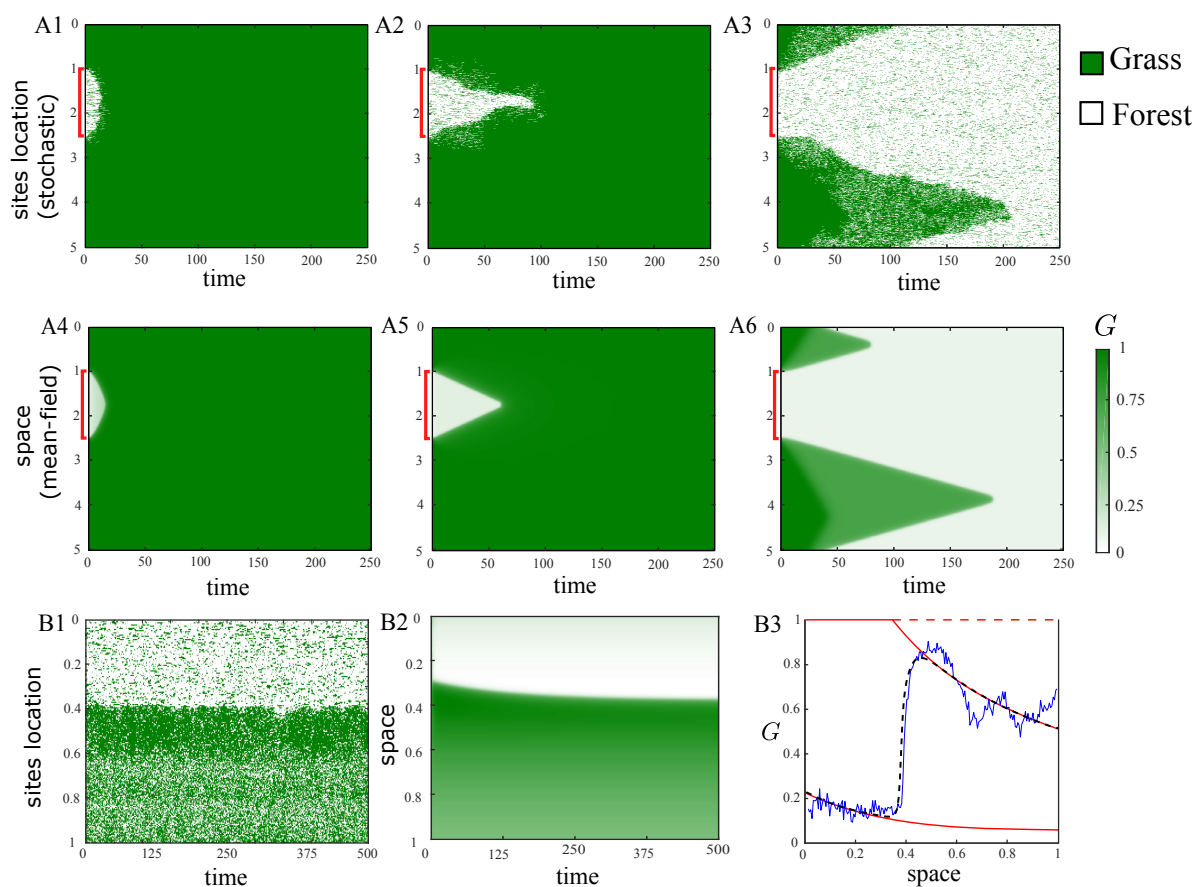


Figure 5. A1–A6: Comparison of solutions of the finite-size stochastic system and the corresponding GKEs for a grass-dominant regime (A1, A2, A4, and A5) and a regime in which forest invades grass (A3 and A6). **Parameters:** $\Gamma = S_5$ the one-dimensional torus of length 5, $N = 3000$, $\bar{J} = 0.5$ (A1, A4), $\bar{J} = 0.9$ (A2, A5), $\bar{J} = 1.25$ (A3, A6), and $\sigma = 0.05$. Initial conditions are the same for all simulations with a block of forest on $[1, 2.5]$ (marked in red). The colorbar for A4–A6 is shown on the right. Sites are either 1 (grass) or 0 (forest) in A1–A3 and hence we use the same colorbar.

B1–B3: Simulated example of front-pinning/Maxwell point phenomenon for the mesoscale model. Panel B1 shows a single realization of the finite-size model, with panel B2 showing the solution to the corresponding GKEs. B3 shows the final solution profile of the GKEs (dashed black line), bifurcation diagram for the GKEs without spatial interaction (stable equilibria in solid red, unstable equilibria in dashed red), and the time averaged profile of the solution to the particle system (solid blue). **Parameters:** $\Gamma = [0, 1]$, $N = 2000$, $\bar{J} = 1.1$, $\sigma = 0.02$ reflecting boundary conditions.

solution profiles of the finite-size model and the GKEs. We observe that the solution of the GKEs (in black) essentially interpolates between the two stable equilibria of the GKEs without spatial interaction (solid red). The solution of the finite-size model approximately matches that of the GKEs but naturally has some stochastic excursions since we are observing a single realization of the process.

4. Proofs of main results. In this section, we provide the details of the proof of Theorem 2.2.1. To this end, we use a bijection between the original state space S^K and the K vectors of the canonical basis in $\mathbb{R}^K$ and reformulate the finite-size Markov process as a stochastic differential equation. This allows a simpler analysis of all models in the class, to prove existence and uniqueness of solutions of the mean-field limit, as well as convergence of the finite system towards the limit system as the system size diverges.

4.1. Mathematical preliminaries. Throughout and in the proofs which follow, we work on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a countable family of i.i.d. Poisson point processes $\mathcal{N} = \{\mathcal{N}_{x,y}^i(t) : (x, y) \in S^K \times S^K, i \in \mathbb{N}, t \geq 0\}$ on $\mathbb{R}_+^2$ with compensators (i.e., predictable part) given by the Lebesgue measure on $\mathbb{R}_+^2$ with $\mathbb{R}_+ := [0, \infty)$. Define the filtration

$$(4.1) \quad \mathcal{F}_t = \sigma(\mathcal{N}_{x,y}^i(A \times B) : (x, y) \in S^K \times S^K, i \in \mathbb{N}, A \in \mathcal{B}(\mathbb{R}_+), B \in \mathcal{B}([0, t])), \quad t \geq 0,$$

where $\mathcal{B}(H)$ denotes the space of Borel sets of H , and $\mathcal{N}_{x,y}^i(A \times B)$ is the number of points of the point process $\mathcal{N}_{x,y}^i$ in $A \times B$. $\mathcal{D}(\mathbb{R}_+; E)$ denotes the Skorohod space of càdlàg functions with values in the space E (typically, considered here to be $\mathbb{R}_+$ or $\mathbb{R}^p$ for some $p \in \mathbb{N}$), i.e., functions that are right-continuous functions with left limits everywhere. We recall the following useful lemma (see Graham and Robert [29]).

Lemma 4.1.1. *Suppose that the processes $Y = \{Y(t) : t \geq 0\}$ and $Z = \{Z(t) : t \geq 0\}$ are in $\mathcal{D}(\mathbb{R}_+; \mathbb{R}_+)$ and are adapted to $(\mathcal{F}_t)_{t \geq 0}$. If $\mathcal{N}$ is a Poisson process on $\mathbb{R}_+^2$ with compensator given by the Lebesgue measure on $\mathbb{R}_+^2$, then the process $I = \{I(t) : t \geq 0\}$ given by*

$$I(t) = \int_0^t Y(s^-) \int_0^\infty \mathbb{1}_{\{0 \leq z \leq Z(s^-)\}} [\mathcal{N}(dz, ds) - dz ds]$$

is a local $\mathcal{F}_t$ -martingale. Moreover, $I(t)$ is a piecewise constant process with jumps of time-dependent size (i.e., $Y(t)$ is the size of a jump at time t) occurring at the times of the jumps of a nonhomogeneous Poisson process with instantaneous rate $Z(t)$.

In particular, under the hypotheses and notation of Lemma 4.1.1, we have

$$\mathbb{E} \left[\int_0^t Y(s^-) \int_0^\infty \mathbb{1}_{\{0 \leq z \leq Z(s^-)\}} \mathcal{N}(dz, ds) \right] = \mathbb{E} \left[\int_0^t Y(s^-) Z(s^-) ds \right],$$

a fact which is used frequently in the arguments which follow.

In order to prove that the mean-field equation introduced in Theorem 2.2.1 admits a well-defined solution, we must consider spatially extended discrete-state continuous-time stochastic processes. Γ always denotes a Borel set in $\mathbb{R}^2$, and q is a probability measure on $(\Gamma, \mathcal{B}(\Gamma))$. To analyze these processes, we work in the space $\mathcal{M}_T$ of stochastic processes $Y = \{Y(t, r) : r \in \Gamma, t \in [0, T]\}$ with state space S^K which are measurable on the product space $(\Omega \times \Gamma)$ and,

for each fixed $r \in \Gamma$, $Y(t, r)$ is $\mathcal{F}_t$ -adapted.⁴ Define the norm $\|\cdot\|_{\mathcal{M}_T}$ on $\mathcal{M}_T$ by

$$(4.2) \quad \|Y\|_{\mathcal{M}_T} = \mathcal{E} \left[\mathbb{E} \left[\sup_{0 \leq s \leq T} \|Y(s, r')\| \right] \right] = \int_{\Gamma} \mathbb{E} \left[\sup_{0 \leq s \leq T} \|Y(s, r')\| \right] dq(r'),$$

where $\|\cdot\|$ denotes the standard Euclidean norm on $\mathbb{R}^K$. The norm $\|\cdot\|_{\mathcal{M}_T}$ identifies processes in $\mathcal{M}_T$ that are $\mathbb{P}$ a.s.- q -a.e. equal. A classical argument shows that the space $\mathcal{M}_T$ is complete under this norm (see, e.g., [4, Chapter 3]).

4.2. Stochastic differential equation formulation of the Markov process. Theorem 2.2.1 deals with an interacting Markov process model with finite state space S^K containing K elements $(x_1, \dots, x_K)$. For convenience and definiteness, we assume (up to bijection) that the elements of S^K are given by the canonical basis of $\mathbb{R}^K$:

$$x_1 := \frac{1}{\sqrt{2}}(1, 0, \dots, 0)^T, \quad x_2 := \frac{1}{\sqrt{2}}(0, 1, 0, \dots, 0)^T, \dots, \quad x_K := \frac{1}{\sqrt{2}}(0, \dots, 0, 1)^T.$$

This choice (and scaling) is particularly convenient in that it allows using the usual Euclidean norm on $\mathbb{R}^K$ as our norm on S^K , and in that norm, all states are equidistant (one unit apart).

Using that representation, the state of site i is a process in $\mathcal{D}(\mathbb{R}_+, S^K)$ whose evolution is described by the Poisson-driven SDE,

$$(4.3) \quad X^i(t) = X^i(0) + \sum_{y \in S^K} \int_0^t (y - X^i(s^-)) \sum_{\substack{x \in S^K, \\ x \neq y}} \mathbb{1}_{\{X^i(s^-)=x\}} \int_0^\infty \mathbb{1}_{\{0 \leq z \leq R_{x,y}^{i,N}(X(s^-))\}} \mathcal{N}_{x,y}^i(dz, ds)$$

for each $t \geq 0$. The transition rate intensity $R_{x,y}^{i,N}$ is given by (2.1), and we write $R_{x,y}^{i,N}(X(s^-))$ to emphasize that the transition rate depends on the entire system state $X(s^-) = \{X^i(s^-) : i \in \{1, \dots, N\}\}$. In (4.3), for each $x, y \in S^K \times S^K$ and $i \in \{1, \dots, N\}$, each $\mathcal{N}_{x,y}^i$ is an independent Poisson point process on $\mathbb{R}_+^2$ with compensator (or predictable part) given by the Lebesgue measure on $\mathbb{R}_+^2$ (see section 4.1). Indeed, the processes

$$\int_0^\infty \mathbb{1}_{\{0 \leq z \leq f(t)\}} \mathcal{N}_{x,y}^i(dz, ds)$$

are inhomogeneous Poisson processes with instantaneous rate $f(t)$, and therefore the system of SDEs defined by equation (4.3) is a Markov process with independent exponentially distributed transitions with rates $R_{x,y}^{i,N}(X(s^-))$.

Under assumptions H1 and H2, we can define the following quantities:

- (i) There exists $L > 0$ such that for all $x, y \in S^K$ and all $X, Y \in R_+$,

$$|\Phi_{x,y}(X) - \Phi_{x,y}(Y)| \leq L |X - Y|.$$

⁴Rigorously, these processes and their measurability are defined through Markov kernels as done in the statement of assumption H_{IC} . Here, measurability is stated to make sure the norm is well defined; the measurability and adaptedness will be obvious for the processes relevant to our proof, and therefore the norm $\|\cdot\|_{\mathcal{M}_T}$ is well defined.

- (ii) $\|\Phi_{x,y}\|_\infty := \sup_{u \in [0, \|W\|_\infty]} |\Phi_{x,y}(u)| < \infty$.
- (iii) $\|W_{x,y}\|_\infty := \sup_{r,r'} |W_{x,y}(r, r')| < \infty$.
- (iv) $\|W\|_\infty := \sup_{x,y \in S^K} \|W_{x,y}\|_\infty < \infty$.

Under assumptions H1, H2, and H_{IC} , strong existence and uniqueness of solutions for the finite-size Markov jump process is classical (see, e.g., [31]).

4.3. The McKean–Vlasov process: Definition, existence, and uniqueness of solutions.

Theorem 2.2.1 states that the finite-size Markov process with transition rates given by (2.1) converges in distribution towards a spatially extended process defined on the support of q on Γ . The limit process, denoted by $\bar{X}$ and referred to as the mean-field limit, has interjump times that are independent and exponentially distributed. The jump rates of $\bar{X}$ at time t depend self-consistently on the law of the process itself and are given by

$$(4.4) \quad x \rightarrow y \quad \Phi_{x,y} \left(\int_{\Gamma} W_{x,y}(r, r') \mathbb{P}[\bar{X}(t^-, r') = \psi(x, y)] dq(r') \right) =: \bar{R}_{x,y}(\bar{X}(t^-), r)$$

for each pair of states $(x, y) \in S^K \times S^K$ and where $\bar{X}(t^-) = \{\bar{X}(t^-, r') : r' \in \Gamma\}$. $\bar{X}$ is non-Markovian because of the dependence of the rates on the law of the solution and not on the state of the system. The mean-field limit described above is a so-called McKean–Vlasov process [58]. Due to the nonstandard nature of these processes, classical existence and uniqueness theory is not sufficient for establishing well-posedness. Although the constructions are standard, we provide the necessary existence and uniqueness arguments to show that the mean-field limit is well-posed in order to keep the presentation self-contained.

Definition 4.3.1. A strong solution of the mean-field process with initial condition $\xi_0(r)$ with a regular distribution in space (in the sense of hypothesis H_{IC}) is a process $\bar{X} \in \mathcal{M}_T$ such that for q -almost every r , $\bar{X}(0, r)$ is $\mathbb{P}$ -almost surely equal to $\xi_0(r)$, and, given the law of the solution up to time t , the process jumps from state x to a state y at a rate $\bar{R}_{x,y}$ given by (4.4).

Equivalently, given a family of Poisson processes $\mathcal{N} = \{\mathcal{N}_{x,y}(t) : (x, y) \in S^K \times S^K, t \geq 0\}$, a solution $\bar{X} \in \mathcal{M}_T$ is a process such that for q -almost every r , $\bar{X}(0, r)$ is $\mathbb{P}$ -almost surely equal to $\xi_0(r)$, and, for q -almost every $r \in \Gamma$ and all $t \in [0, T]$, the following equation holds $\mathbb{P}$ -almost surely:

$$(4.5) \quad \bar{X}(t, r) = \xi_0(r) + \sum_{y \in S^K} \int_0^t (y - \bar{X}(s^-, r)) \sum_{\substack{x \in S^K, \\ x \neq y}} \mathbb{1}_{\{\bar{X}(s^-, r) = x\}} \int_0^\infty \mathbb{1}_{\{0 \leq z \leq \bar{R}_{x,y}(\bar{X}(s^-, r))\}} \mathcal{N}_{x,y}(dz, ds).$$

The equivalence between the process associated with transition rates (4.4) and the process which solves the mean-field equation (4.5) follows from the same classical arguments as those allowing us to write up the Poisson driven SDEs describing the evolution of the finite-size system. Indeed, if the solution to (4.5) exists, and assuming that $\bar{X}(t^-, r) = x$, then a jump of size $(y - x)$ (therefore, moving the process from x to $y = (y - x) + x$) arises at a random, exponentially distributed time with rate $\bar{R}_{x,y}(\bar{X}(t^-), r)$.

Definition 4.3.2. The solution to (4.5) is unique if for any two strong solutions $\bar{X}$ and $\tilde{X}$ to (4.5), the event

$$\left\{ \bar{X}(t, r) = \tilde{X}(t, r) \text{ for each } t \in [0, T] \right\}$$

has probability 1 with respect to the product measure $\mathbb{P} \otimes q$.

Theorem 4.3.3. Suppose H1 and H2 hold. Let $\xi_0(r)$ be a spatially measurable random variable on S^K (in the sense of H_{IC}) independent of the Poisson processes $\mathcal{N}_{x,y}$ for $x, y \in S^K$. The mean-field equation (4.5) with initial condition $\xi_0(r)$ has a unique strong solution in $\mathcal{M}_T$.

Proof. Define the mapping Ψ which acts on processes $X = \{X(t, r) : r \in \Gamma, t \in [0, T]\}$ in $\mathcal{M}_T$ according to

$$\begin{aligned} \Psi(X)(t, r) = & \xi_0(r) \\ & + \sum_{y \in S^K} \int_0^t (y - X(s^-, r)) \sum_{\substack{x \in S^K, \\ x \neq y}} \mathbb{1}_{\{X(s^-, r) = x\}} \int_0^\infty \mathbb{1}_{\{0 \leq z \leq \bar{R}_{x,y}(X(s^-, r))\}} \mathcal{N}_{x,y}(dz, ds) \end{aligned}$$

for each $(t, r) \in \mathbb{R}_+ \times \Gamma$.

According to Definition 4.3.1, a solution to the mean-field equation (4.5) is a fixed point of Ψ . Therefore, proving the theorem amounts to showing (i) Ψ has a fixed point, and (ii) the solution obtained as the fixed point of Ψ is unique according to Definition 4.3.2. Because of the consistency between Definition 4.3.2 and the norm $\|\cdot\|_{\mathcal{M}_T}$ given by (4.2), uniqueness will follow if all solutions are indistinguishable under $\|\cdot\|_{\mathcal{M}_T}$. We demonstrate both properties using a classical argument based on Picard's iterates. In order to be able to iterate the map Ψ , we must first establish that for any $X \in \mathcal{M}_T$, the process $Y := \Psi(X)$ also belongs to $\mathcal{M}_T$. For $X \in \mathcal{M}_T$, the integrals over Γ in the jump intensity functionals

$$\bar{R}_{x,y}(X(s^-), r) = \Phi_{x,y} \left(\int_{\Gamma} W_{x,y}(r, r') \mathbb{P}[X(s^-, r') = \psi(x, y)] dq(r') \right), \quad y \in S^K,$$

are $\mathcal{F}_s$ -measurable, continuous with respect to r , and finite due to the $(\Omega \times \Gamma)$ -measurability and boundedness of X , and to the regularity and boundedness of the $W_{x,y}$'s and $\Phi_{x,y}$'s, respectively (see H1 and H2). Therefore, each $Y(t, r)$ is $\mathcal{F}_t$ -adapted and measurable with respect to $(\Omega \times \Gamma)$. Since $X(t, r) \in S^K$, $Y(t, r)$ takes values in S^K as well. Therefore, Ψ maps $\mathcal{M}_T$ to itself. We can thus define the sequence of processes $(X^k(r))_{k \geq 0}$ as follows:

$$X^0 = \{X^0(t, r) : X^0(t, r) = \xi_0(r), r \in \Gamma, t \in [0, T]\},$$

with the S^K -valued random variable $\xi_0(r)$ chosen to be $(\Omega \times \Gamma)$ -measurable and

$$X^{k+1} = \left\{ X^{k+1}(t, r) : X^{k+1}(t, r) = \Psi(X^k)(t, r), r \in \Gamma, t \in [0, T] \right\}, \quad k \geq 1.$$

For $k \geq 1$, standard estimation of the quantity $\|X^{k+1} - X^k\|_{\mathcal{M}_T}$ using Lemma 4.1.1 yields

$$\begin{aligned} \|X^{k+1} - X^k\|_{\mathcal{M}_T} &= \mathcal{E} \left[\mathbb{E} \left[\sup_{0 \leq t \leq T} \|X^{k+1}(t, r) - X^k(t, r)\| \right] \right] \\ &\leq \sum_{y \in S^K} \sum_{\substack{x \in S^K, \\ x \neq y}} \left\{ \hat{A}_{x,y}(T) + \hat{B}_{x,y}(T) + \hat{C}_{x,y}(T) \right\}, \end{aligned}$$

where

$$\begin{aligned} \hat{A}_{x,y}(T) &:= \mathcal{E} \left[\mathbb{E} \left[\int_0^T \|y - X^k(s^-, r)\| \left| \bar{R}_{x,y}(X^k(s^-), r) - \bar{R}_{x,y}(X^{k-1}(s^-), r) \right| ds \right] \right], \\ \hat{B}_{x,y}(T) &:= \mathcal{E} \left[\mathbb{E} \left[\int_0^T \|X^k(s^-, r) - X^{k-1}(s^-, r)\| \left| \bar{R}_{x,y}(X^{k-1}(s^-), r) \right| ds \right] \right], \end{aligned}$$

and

$$\hat{C}_{x,y}(T) := \mathcal{E} \left[\mathbb{E} \left[\int_0^T \left| \mathbb{1}_{\{X^k(s^-, r)=x\}} - \mathbb{1}_{\{X^{k-1}(s^-, r)=x\}} \right| \left| \bar{R}_{x,y}(X^{k-1}(s^-), r) \right| ds \right] \right].$$

Since $\|y - X^k(s^-, r)\| \leq 1$,

$$(4.6) \quad \hat{A}_{x,y}(T) \leq \mathcal{E} \left[\mathbb{E} \left[\int_0^T \left| \bar{R}_{x,y}(X^k(s^-), r) - \bar{R}_{x,y}(X^{k-1}(s^-), r) \right| ds \right] \right].$$

Estimating the integrand in (4.6) using the Lipschitz continuity of $\Phi_{x,y}$ and the boundedness of $W_{x,y}$ yields

$$\begin{aligned} &\left| \bar{R}_{x,y}(X^k(s^-), r) - \bar{R}_{x,y}(X^{k-1}(s^-), r) \right| \\ &\leq L \|W_{x,y}\|_{\infty} \int_{\Gamma} \left| \mathbb{P} \left[X^k(s^-, r') = \psi(x, y) \right] - \mathbb{P} \left[X^{k-1}(s^-, r') = \psi(x, y) \right] \right| dq(r'). \end{aligned}$$

For each fixed r' and each fixed $s \geq 0$, we have

$$\begin{aligned} &\left| \mathbb{P} \left[X^k(s^-, r') = \psi(x, y) \right] - \mathbb{P} \left[X^{k-1}(s^-, r') = \psi(x, y) \right] \right| \\ &\leq \mathbb{E} \left[\left| \mathbb{1}_{\{X^k(s^-, r')=\psi(x,y)\}} - \mathbb{1}_{\{X^{k-1}(s^-, r')=\psi(x,y)\}} \right| \right] \\ &\leq \mathbb{E} \left[\|X^k(s^-, r') - X^{k-1}(s^-, r')\| \right], \end{aligned}$$

since

$$\|X^k(s^-, r') - X^{k-1}(s^-, r')\| = \begin{cases} 0, & X^k(s^-, r') = X^{k-1}(s^-, r'), \\ 1, & X^k(s^-, r') \neq X^{k-1}(s^-, r'). \end{cases}$$

Thus

$$\begin{aligned} & \left| \bar{R}_{x,y} \left(X^k(s^-), r \right) - \bar{R}_{x,y} \left(X^{k-1}(s^-), r \right) \right| \\ & \leq L \|W_{x,y}\|_\infty \int_\Gamma \mathbb{E} \left[\|X^k(s^-, r') - X^{k-1}(s^-, r')\| \right] dq(r') \\ & \leq L \|W_{x,y}\|_\infty \left\| X^k - X^{k-1} \right\|_{\mathcal{M}_s}. \end{aligned}$$

Returning to (4.6) and using the estimate above gives

$$\hat{A}_{x,y}(T) \leq L \|W_{x,y}\|_\infty \int_0^T \left\| X^k - X^{k-1} \right\|_{\mathcal{M}_s} ds.$$

The estimation for $\hat{B}_{x,y}$ follows simply from using the boundedness of $\Phi_{x,y}$ to obtain

$$\hat{B}_{x,y}(T) \leq \|\Phi_{x,y}\|_\infty \int_0^T \mathcal{E} \left[\mathbb{E} \left[\|X^k(s^-, r) - X^{k-1}(s^-, r)\| \right] \right] ds.$$

Similarly, $\hat{C}_{x,y}$ can be bounded as follows:

$$\begin{aligned} \hat{C}_{x,y}(T) & \leq \mathcal{E} \left[\mathbb{E} \left[\int_0^T \left| \mathbb{1}_{\{X^k(s^-, r)=x\}} - \mathbb{1}_{\{X^{k-1}(s^-, r)=x\}} \right| \left| \bar{R}_{x,y} \left(X^{k-1}(s^-), r \right) \right| ds \right] \right] \\ & \leq \|\Phi_{x,y}\|_\infty \int_0^T \mathcal{E} \left[\mathbb{E} \left[\left| \mathbb{1}_{\{X^k(s^-, r)=x\}} - \mathbb{1}_{\{X^{k-1}(s^-, r)=x\}} \right| \right] \right] ds \\ & \leq \|\Phi_{x,y}\|_\infty \int_0^T \mathcal{E} \left[\mathbb{E} \left[\left\| X^k(s^-, r) - X^{k-1}(s^-, r) \right\| \right] \right] ds \\ & \leq \|\Phi_{x,y}\|_\infty \int_0^T \left\| X^k - X^{k-1} \right\|_{\mathcal{M}_s} ds. \end{aligned}$$

Combining the estimates for $\hat{A}_{x,y}$, $\hat{B}_{x,y}$, and $\hat{C}_{x,y}$ yields

$$\left\| X^{k+1} - X^k \right\|_{\mathcal{M}_T} \leq \bar{K} \int_0^T \left\| X^k - X^{k-1} \right\|_{\mathcal{M}_s} ds \quad \text{for each } k \geq 1,$$

where

$$\bar{K} = \sum_{y \in S^K} \sum_{\substack{x \in S^K, \\ x \neq y}} (L \|W_{x,y}\|_\infty + 2 \|\Phi_{x,y}\|_\infty).$$

Therefore

$$\left\| X^{k+1} - X^k \right\|_{\mathcal{M}_T} \leq \frac{(\bar{K}T)^k}{k!} \left\| X^1 - X^0 \right\|_{\mathcal{M}_T} \leq \frac{(\bar{K}T)^k}{k!}, \quad k \geq 1.$$

Thus $(X^k)_{k \geq 0}$ is a Cauchy sequence in the complete space $\mathcal{M}_T$ and converges to a limit in $\mathcal{M}_T$. The conclusion of the theorem is standard at this point (see, e.g., [43]) and yields the existence of an $\mathcal{F}_t$ -adapted $(\Omega \times \Gamma)$ -measurable process $\bar{X} = \{\bar{X}(t, r) : r \in \Gamma, t \in [0, T]\} \in \mathcal{M}_T$ such that $\bar{X} = \Psi(\bar{X})$. Therefore $\bar{X}$ is a strong solution to (4.5) on $[0, T]$.

The estimates above can be used to show that for any two solutions $\bar{X}$ and $\tilde{X}$ to (4.5),

$$\|\bar{X} - \tilde{X}\|_{\mathcal{M}_T} \leq K' \int_0^T \|\bar{X} - \tilde{X}\|_{\mathcal{M}_s} ds.$$

Using Gronwall's lemma and the fact that the two solutions have identical initial conditions allows us to conclude that $\bar{X} = \tilde{X}$ in the norm $\|\cdot\|_{\mathcal{M}_T}$, and hence solutions to (4.5) are unique in the sense of Definition 4.3.2. ■

4.4. Convergence towards the McKean–Vlasov equation. We now undertake the proof of the convergence result in Theorem 2.2.1.

Proof of Theorem 2.2.1.

I. *Convergence.* Fix $N \in \mathbb{N}$, $i \in \{1, \dots, N\}$, a time $\tau > 0$, and a fixed configuration of sites $(r_1, \dots, r_N)$ drawn as i.i.d. random variables according to the probability measure q on Γ . We proceed to demonstrate that the process X which solves (4.3) converges almost surely with respect to $\mathbb{P} \otimes q$ to a *coupled* process $\bar{X}^i$, a particular solution of (4.5) with the same initial condition and Poisson processes as site i :

$$\begin{aligned} \bar{X}^i(t, r) &= \xi_0^i(r) \\ &+ \sum_{y \in S^K} \int_0^t (y - \bar{X}^i(s^-, r)) \sum_{\substack{x \in S^K, \\ x \neq y}} \mathbb{1}_{\{\bar{X}^i(s^-, r) = x\}} \int_0^\infty \mathbb{1}_{\{0 \leq z \leq \bar{R}_{x,y}(\bar{X}^i(s^-), r)\}} \mathcal{N}_{x,y}^i(dz, ds). \end{aligned}$$

Estimate the distance between X^i and $\bar{X}^i$ in the $\|\cdot\|_{\mathcal{M}_T}$ norm. Because initial conditions are identical, they disappear in the norm, and because the Poisson processes are also identical, the stochastic terms combine and we obtain

$$(4.7) \quad \mathbb{E} \left[\sup_{0 \leq t \leq \tau} \|X^i(t) - \bar{X}^i(t, r_i)\| \right] \leq \sum_{y \in S^K} \sum_{\substack{x \in S^K, \\ x \neq y}} A_{x,y}(\tau) + B_{x,y}(\tau) + C_{x,y}(\tau),$$

where

$$\begin{aligned} A_{x,y}(\tau) &= \mathbb{E} \left[\int_0^\tau \|y - X^i(s)\| \left| \mathbb{1}_{\{X^i(s^-) = x\}} - \mathbb{1}_{\{\bar{X}^i(s^-, r_i) = x\}} \right| |R_{x,y}^{i,N}(X(s^-))| ds \right], \\ B_{x,y}(\tau) &= \mathbb{E} \left[\int_0^\tau \|y - X^i(s^-)\| |\bar{R}_{x,y}(\bar{X}^i(s^-), r_i) - R_{x,y}^{i,N}(X(s^-))| ds \right], \\ C_{x,y}(\tau) &= \mathbb{E} \left[\int_0^\tau \|\bar{X}^i(s^-, r_i) - X^i(s^-)\| \left| \mathbb{1}_{\{\bar{X}^i(s^-, r_i) = x\}} \right| |\bar{R}_{x,y}(\bar{X}^i(s^-), r_i)| ds \right]. \end{aligned}$$

We immediately obtain an upper bound on $A_{x,y}$ since $\|y - X^i(s)\| \leq 1$ and $R_{x,y}^{i,N}$ is bounded due to the boundedness of each $\Phi_{x,y}$. In particular,

$$\begin{aligned} \mathcal{E}[A_{x,y}(\tau)] &\leq \|\Phi_{x,y}\|_\infty \mathcal{E} \left[\mathbb{E} \left[\int_0^\tau \left| \mathbb{1}_{\{X^i(s^-) = x\}} - \mathbb{1}_{\{\bar{X}^i(s^-, r_i) = x\}} \right| ds \right] \right] \\ &\leq \|\Phi_{x,y}\|_\infty \int_0^\tau \mathcal{E} \left[\mathbb{E} \left[\sup_{0 \leq u \leq s} \|X^i(u^-) - \bar{X}^i(u^-, r_i)\| \right] \right] ds. \end{aligned}$$

Similarly, it is straightforward to derive the following estimate on $C_{x,y}$:

$$\mathcal{E}[C_{x,y}(\tau)] \leq \|\Phi_{x,y}\|_\infty \int_0^\tau \mathcal{E} \left[\mathbb{E} \left[\sup_{0 \leq u \leq s} \|\bar{X}^i(u^-, r_i) - X^i(u)\| \right] \right] ds.$$

The requisite estimation for $B_{x,y}$ is nontrivial; we claim that

$$(4.8) \quad \mathcal{E}[B_{x,y}(\tau)] \leq L \|W_{x,y}\|_\infty \int_0^\tau \mathcal{E} \left[\mathbb{E} \left[\sup_{0 \leq u \leq s} \|\bar{X}^i(u, r_i) - X^i(u)\| \right] \right] ds + \tau L \sqrt{\frac{C}{N}}$$

for some constant $C > 0$ which is independent of X^i , $\bar{X}^i$, and $\mathcal{A}_N$. To establish that (4.8) holds, begin by estimating as follows:

$$(4.9) \quad \mathcal{E}[B_{x,y}(\tau)] \leq \int_0^\tau \mathcal{E} \left[\mathbb{E} \left[\|\bar{R}_{x,y}(\bar{X}^i(s^-), r_i) - R_{x,y}^{i,N}(X(s^-))\| \right] \right] ds.$$

Next make the following estimate on the difference of the transition rate functions using the Lipschitz continuity of $\Phi_{x,y}$:

$$\begin{aligned} & |\bar{R}_{x,y}(\bar{X}^i(s^-), r_i) - R_{x,y}^{i,N}(X(s^-))| \\ & \leq L \left| \frac{1}{N} \sum_{j=1}^N W_{x,y}(r_i, r_j) \mathbb{1}_{\{X^j(s^-) = \psi(x,y)\}} - \int_\Gamma W_{x,y}(r_i, r') \mathbb{P}[\bar{X}^i(s^-, r') = \psi(x,y)] dq(r') \right|. \end{aligned}$$

Thus

$$\begin{aligned} & |\bar{R}_{x,y}(\bar{X}^i(s^-), r_i) - R_{x,y}^{i,N}(X(s^-))| \\ & \leq \frac{L}{N} \sum_{j=1}^N W_{x,y}(r_i, r_j) \left| \mathbb{1}_{\{X^j(s^-) = \psi(x,y)\}} - \mathbb{1}_{\{\bar{X}^i(s^-, r_j) = \psi(x,y)\}} \right| \\ & \quad + L \left| \frac{1}{N} \sum_{j=1}^N W_{x,y}(r_i, r_j) \mathbb{1}_{\{\bar{X}^i(s^-, r_j) = \psi(x,y)\}} - \int_\Gamma W_{x,y}(r_i, r') \mathbb{P}[\bar{X}^i(s^-, r') = \psi(x,y)] dq(r') \right| \\ & =: D_{x,y}^1(s) + D_{x,y}^2(s). \end{aligned}$$

The first term above is easily estimated using the boundedness of the kernels $W_{x,y}$,

$$\begin{aligned} & \mathcal{E}[\mathbb{E}[D_{x,y}^1(s)]] \\ & \leq \frac{L \|W_{x,y}\|_\infty}{N} \sum_{j=1}^N \mathcal{E} \left[\mathbb{E} \left[\left| \mathbb{1}_{\{X^j(s^-) = \psi(x,y)\}} - \mathbb{1}_{\{\bar{X}^i(s^-, r_j) = \psi(x,y)\}} \right| \right] \right] \\ & \leq \frac{L \|W_{x,y}\|_\infty}{N} \sum_{j=1}^N \mathcal{E} \left[\mathbb{E} \left[\left| X^j(s^-) - \bar{X}^i(s^-, r_j) \right| \right] \right]. \end{aligned}$$

For each $j \in \{1, \dots, N\}$ and each fixed $s \geq 0$, the random variables $\|X^j(s^-) - \bar{X}^i(s^-, r_j)\|$ are identically distributed given the configuration $\mathcal{A}_N$. In other words, the quantity

$$\mathcal{E} \left[\mathbb{E} \left[\|X^j(s^-) - \bar{X}^i(s^-, r_j)\| \right] \right]$$

does not depend on j . Hence

$$\mathcal{E} \left[\mathbb{E} \left[D_{x,y}^1(s) \right] \right] \leq L \|W_{x,y}\|_\infty \mathcal{E} \left[\mathbb{E} \left[\|X^k(s^-) - \bar{X}^i(s^-, r_k)\| \right] \right]$$

for each $k \in \{1, \dots, N\}$. It follows readily that

$$(4.10) \quad \int_0^\tau \mathcal{E} \left[\mathbb{E} \left[D_{x,y}^1(s) \right] \right] ds \leq L \|W_{x,y}\|_\infty \int_0^\tau \mathcal{E} \left[\mathbb{E} \left[\sup_{0 \leq u \leq s} \|X^k(u^-) - \bar{X}^i(u^-, r_k)\| \right] \right] ds$$

for each $k \in \{1, \dots, N\}$.

In order to estimate $D_{x,y}^2(s)$, note that the collection of random variables

$$\left(W_{x,y}(r_i, r_j) \mathbb{1}_{\{\bar{X}^i(s^-, r_j) = \psi(x, y)\}} \right)_{j \in \{1, \dots, N\}}$$

are conditionally i.i.d. given r_i . Hence the expression

$$(4.11) \quad \frac{1}{N} \sum_{j=1}^N W_{x,y}(r_i, r_j) \mathbb{1}_{\{\bar{X}^i(s^-, r_j) = \psi(x, y)\}}$$

is, conditionally on r_i , the sum of i.i.d. random variables with finite mean and variance (since the process $\bar{X}^i$ is a.s. bounded). It is then natural to define $\hat{\mathbb{E}}_i$, the conditional expectation operator on $\Omega' \times \Omega$ given r_i . The mean of each summand in (4.11) with respect to $\hat{\mathbb{E}}_i$ is given by

$$(4.12) \quad \hat{\mathbb{E}}_i \left[W_{x,y}(r_i, r_j) \mathbb{1}_{\{\bar{X}^i(s^-, r_j) = \psi(x, y)\}} \right] = \int_\Gamma W_{x,y}(r_i, r') \mathbb{P} [\bar{X}^i(s^-, r') = \psi(x, y)] dq(r').$$

Let

$$F_{x,y}^1(s) = \frac{1}{N} \sum_{j=1}^N W_{x,y}(r_i, r_j) \mathbb{1}_{\{\bar{X}^i(s^-, r_j) = \psi(x, y)\}}$$

and

$$F_{x,y}^2(s) = \int_\Gamma W_{x,y}(r_i, r') \mathbb{P} [\bar{X}^i(s, r') = \psi(x, y)] dq(r').$$

Thus, for each fixed $s \in [0, \tau]$, Hölder's inequality yields

$$(4.13) \quad \begin{aligned} \hat{\mathbb{E}}_i \left[|F_{x,y}^1(s) - F_{x,y}^2(s)| \right] &\leq \left(\hat{\mathbb{E}}_i \left[(F_{x,y}^1(s) - F_{x,y}^2(s))^2 \right] \right)^{1/2} \\ &\leq \frac{1}{N} \left(\sum_{j=1}^N \hat{\mathbb{E}}_i \left[\left(W_{x,y}(r_i, r_j) \mathbb{1}_{\{\bar{X}^i(s^-, r_j) = \psi(x, y)\}} - F_{x,y}^2(s) \right)^2 \right] \right)^{1/2}, \end{aligned}$$

where we have used that

$$\text{Cov} \left(W_{x,y}(r_i, r_j) \mathbb{1}_{\{\bar{X}^i(s^-, r_j) = \psi(x,y)\}}, W_{x,y}(r_i, r_k) \mathbb{1}_{\{\bar{X}^i(s^-, r_k) = \psi(x,y)\}} \right) = 0$$

for $j \neq k$, conditional on r_i . By (4.12), the summands in (4.13) are variances of (conditionally) i.i.d. random variables. Thus

$$\hat{\mathbb{E}}_i [|F_{x,y}^1(s) - F_{x,y}^2(s)|] \leq \frac{1}{\sqrt{N}} \left(\hat{\mathbb{E}}_i \left[\left(W_{x,y}(r_i, r_k) \mathbb{1}_{\{\bar{X}^i(s^-, r_k) = \psi(x,y)\}} - F_{x,y}^2(s) \right)^2 \right] \right)^{1/2}$$

for any $k \neq i$. Since $\bar{X}^i$ is a bounded process, we can uniformly bound its variance by a deterministic constant $C > 0$. Thus

$$\hat{\mathbb{E}}_i [|F_{x,y}^1(s) - F_{x,y}^2(s)|] \leq \sqrt{\frac{C}{N}} \quad \text{for some } C > 0.$$

Therefore, by the tower property of expectation,

$$\mathcal{E} [\mathbb{E} [|F_{x,y}^1(s) - F_{x,y}^2(s)|]] \leq \sqrt{\frac{C}{N}}.$$

Finally, integrate over $[0, T]$ to obtain the desired estimate for $D_{x,y}^2$, i.e.,

$$\begin{aligned} \int_0^\tau \mathcal{E} [\mathbb{E} [|D_{x,y}^2(s)|]] ds &\leq L \int_0^\tau \mathcal{E} [\mathbb{E} [|F_{x,y}^1(s) - F_{x,y}^2(s)|]] ds \\ &\leq \tau L \sqrt{\frac{C}{N}}. \end{aligned}$$

This estimate establishes the claimed inequality for $B_{x,y}$ from (4.8). Finally, return to (4.7) and apply the estimates on $A_{x,y}$, $B_{x,y}$, and $C_{x,y}$ derived above to conclude that

$$\begin{aligned} \mathcal{E} \left[\mathbb{E} \left[\sup_{0 \leq t \leq \tau} ||X^i(t) - \bar{X}^i(t, r_i)|| \right] \right] &\leq K_1 \int_0^\tau \mathcal{E} \left[\mathbb{E} \left[\sup_{0 \leq u \leq s} ||X^i(s) - \bar{X}^i(s, r_i)|| \right] \right] ds \\ &\quad + \frac{K_2}{\sqrt{N}}, \end{aligned}$$

where

$$K_1 := \sum_{y \in S^K} \sum_{\substack{x \in S^K \\ x \neq y}} 2||\Phi_{x,y}||_\infty + L ||W_{x,y}||_\infty$$

and

$$K_2 = \tau \sqrt{C} L K (K - 1) > 0.$$

Gronwall's inequality now shows that

$$\mathcal{E} \left[\mathbb{E} \left[\sup_{0 \leq t \leq \tau} ||X^i(t) - \bar{X}^i(t, r_i)|| \right] \right] \leq \frac{K_1 K_2 \tau}{\sqrt{N}} e^{K_1 \tau} + \frac{K_2 \tau}{\sqrt{N}},$$

and letting $N \rightarrow \infty$ shows that $X^i(t)$ converges almost surely towards $\bar{X}^i(t, r_i)$, which in particular implies the convergence in law stated in the theorem. ■

4.5. Generalized Kolmogorov equations. We now prove Theorem 2.2.1 II, the characterization of the law of the mean-field limit through the so-called generalized Kolmogorov equations (GKEs). Classical Kolmogorov equations are linear differential equations governing the evolution in time of the probability to be at a given state for a continuous-time Markov process. The mean-field limit is non-Markovian, as its transition rates depend on the distribution of the solution at all other locations. In section 4.3 we proved that there exists a unique strong solution to the mean-field process. Therefore, the rates of transition between different states given by (4.4) are well defined, since they depend only on that probability distribution, which is in turn well defined. Let $\bar{X}(t, r)$ be a solution of the mean-field process, recall that $P_x(t, r) = \mathbb{P}[\bar{X}(t, r) = x]$, and define for each pair of distinct states $x, y \in S^K$ the functions

$$\Lambda_{x,y}(t) = \Phi_{x,y} \left(\int_{\Gamma} W_{x,y}(r, r') P_{\psi(x,y)}(t^-, r') dq(r') \right).$$

These functions are well defined, nonnegative, and bounded, and define univocally an auxiliary time inhomogeneous Markov process $Y(t)$ with transition rate from state x to state $y \neq x$ at time t given by $\Lambda_{x,y}(t)$. Standard theory from Markov processes ensures that the law of process $Y(t)$, denoted by Q , satisfies the usual Kolmogorov equations (KE). In detail, the transition probabilities

$$Q_{x,y}(t) = \mathbb{P}[Y(t) = y \mid X(0) = x]$$

satisfy the differential equations

$$\begin{cases} \frac{d}{dt} Q_{x,y}(t) = \sum_{z \neq y} Q_{x,z}(t) \Lambda_{z,y}(t) - \Lambda_y(t) Q_{x,y}(t) & \text{(forward KEs),} \\ \frac{d}{dt} Q_{x,y}(t) = \sum_{z \neq x} \Lambda_{x,z}(t) Q_{z,y}(t) - \Lambda_x(t) Q_{x,y}(t) & \text{(backward KEs),} \end{cases}$$

with $\Lambda_x(t) := \sum_{z \neq x} \Lambda_{x,z}(t)$ for each $t \geq 0$.

The probability distribution of $Y(t)$ is equal to that of the non-Markovian McKean–Vlasov process (4.5) because the Markov process Y has (by definition) independent exponentially distributed transitions with rates $\Lambda_{x,y}(t)$, exactly as the McKean–Vlasov process $\bar{X}$. Therefore, the transition probabilities

$$P_{x,y}(t, r) = \mathbb{P}[\bar{X}(t, r) = y \mid \bar{X}(0, r) = x]$$

are equal to $Q_{x,y}(t)$. The linear KEs for the transition probabilities of Y thus convert into the following nonlinear equations for those of $\bar{X}$:

$$\begin{cases} \partial_t P_{x,y}(t, r) = \sum_{z \neq y} P_{x,z}(t, r) \Lambda_{z,y}(t) - \Lambda_y(t) P_{x,y}(t, r) & \text{(forward GKEs),} \\ \partial_t P_{x,y}(t, r) = \sum_{z \neq x} \Lambda_{x,z}(t) P_{z,y}(t, r) - \Lambda_x(t) P_{x,y}(t, r) & \text{(backward GKEs).} \end{cases}$$

Given that $P_x(t, r) = \sum_y P_{y,x}(t, r) \mu_0(r, y)$ with $\mu_0(r, y) = \mathbb{P}[\bar{X}(r, 0) = y]$, we thus obtain

$$(4.14) \quad \partial_t P_x(t, r) = \sum_{z \neq x} P_z(t, r) \Lambda_{z,x}(t) - \Lambda_x(t) P_x(t, r), \quad x \in S^K,$$

using the forward GKEs above. The system of equations given by (4.14) is precisely the self-consistent K -dimensional nonlinear system of IDEs stated in the theorem.

4.6. Propagation of chaos. We complete the proof of Theorem 2.2.1 by proving the propagation of chaos property. In section 4.4, we showed an almost sure convergence of X^i towards the coupled mean-field process $\bar{X}^i(\cdot, r_i)$ introduced in the proof. For $i \neq j$, because the initial conditions $\xi^i(r)$ and $\xi^j(r)$ are independent (assumption H_{IC}), and because for any (x, x', y, y') elements of S^K , the Poisson processes $\mathcal{N}_{x,y}^i$ and $\mathcal{N}_{x',y'}^j$ are independent, the processes $\bar{X}^i(\cdot, r_i)$ and $\bar{X}^j(\cdot, r_j)$ are independent. Similarly, for any sequence of p distinct indices $(i_1, \dots, i_p)$, the processes $\bar{X}^{i_1}(t, r_{i_1}), \dots, \bar{X}^{i_p}(t, r_{i_p})$ are mutually independent, and the collection of processes $(X^{i_1}, \dots, X^{i_p})$ converges almost surely towards $(\bar{X}^{i_1}(\cdot, r_{i_1}), \dots, \bar{X}^{i_p}(\cdot, r_{i_p}))$, completing the proof.

Appendix A. Numerical methods and codes. In this appendix we describe in more detail the theory and numerical methods behind the figures in the main text. All code needed to reproduce the figures is available on github.com/Touboul-Lab and was run on MATLAB version R2019B. In the figures where the bifurcation diagram of an ODE was overlaid on the result of a simulation, the bifurcation diagram was computed in AUTO [23], and the axes were aligned in Inkscape.

A.1. Numerical parameters. For all numerical calculations and simulations, we take the following smooth functional forms for the fire threshold functions ϕ and ω :

$$\omega(x) = \omega_0 + \frac{\omega_1 - \omega_0}{1 + e^{-(x-\theta_1)/s_1}}, \quad \phi(x) = \phi_0 + \frac{\phi_1 - \phi_0}{1 + e^{-(x-\theta_2)/s_2}} \quad \text{for } x \in [0, 1],$$

with parameter values as given in Table 2.

Table 2

Parameter	μ	ν	ω_0	ω_1	t_1	s_1	ϕ_0	ϕ_1	t_2	s_2
Value	0.1	0.05	0.9	0.4	0.4	0.01	0.1	0.9	0.4	0.05

A.2. Figure 2. Figure 2 shows the results of simulations of the macroscale interacting particle system model for different values of the seed dispersal intensity $\bar{J}$ across a range of initial conditions. There is a single patch $M = 1$ and $N = 3000$ sites in each simulation, and we used the Gillespie algorithm to simulate transition times. For each value of $\bar{J}$ shown, we ran 20 simulations with the initial grass/forest proportions randomly assigned with a weighting which we varied from 0 to 1 to observe a full range of initial conditions. Each blue starred point is the proportion of grass in the particle system at time $t = 100$. The red curves are the bifurcation diagram in $\bar{J}$ of the ODE, i.e., the GKEs of the mean-field limit,

$$\frac{d}{dt}P_G(t) = (1 - P_G(t))\phi(P_G(t)) - \bar{J}P_G(t)(1 - P_G(t))$$

with $P_G(t)$ the probability of any given site in the patch to be covered by grass at time t . Solid red lines are stable equilibrium curves, dashed red lines denote unstable equilibria, and solid red dots mark the bifurcation points.

When multiple distinct states were observed in the particle system at time $t = 100$, we recorded the proportions of grass in the initial conditions and plotted with a solid blue

dot the lowest proportion whose corresponding simulation ended up on the upper branch of the bifurcation diagram. Similarly, the highest initial proportion which resulted in a time $t = 100$ state on the lower branch is recorded with a solid green dot. Tracking which initial conditions result in which end states gives an idea of the “basin of attraction” for the transient states of the particle system—we see a close correspondence between this informal “stochastic attraction basin” and the basin of attraction of the stable states of the GKEs.

A.3. Figure 3. Panels A1 and A2 are obtained by directly computing the principal eigenvalue and corresponding normalized eigenvector of the transition intensity matrix Q of the irreducible part of the Markov chain (i.e., the finite-size model).

A.4. Figure 4. Figure 4 shows a comparison between the macroscale particle system model with $M = 1$ patch and $N = 3000$ sites, and the corresponding GKEs of the mean-field limit, i.e.,

$$\begin{aligned}\dot{G} &= \mu S + \nu T + \phi(G)F - \bar{J}GF - \beta GT, \\ \dot{S} &= -\mu S - \omega(G)S - \bar{J}SF + \beta GT, \\ \dot{T} &= -\nu T + \omega(G)S - \bar{J}TF, \\ \dot{F} &= \bar{J}(G + S + T)F - \phi(G)F, \\ 1 &= G + S + T + F.\end{aligned}$$

In panel A1, we show the solution to the GKEs versus time—the GKEs are a system of ODEs and are solved using an explicit Euler method with step size 0.01; parameters are as in Table 2 with $\bar{J} = 0.25$ and $\beta = 0.4$. Panel C1 also shows solutions of the GKEs solved via the Euler method, but now $\beta = 0.4$ is fixed while $\bar{J}$ is varied to show solutions approaching a heteroclinic cycle in the phase space.

Panels A2, B1, and C2 show the results of direct simulations of the four-species macroscale particle system using the Gillespie algorithm to simulate transition times. We have a single patch $M = 1$ and $N = 3000$ sites with $\bar{J}$ and β as in panel A1. Panel A2 shows the evolution of the proportions of sites in each state, while panel B1 shows the full solution with the state of each site at each time recorded—the sites are essentially in a synchronized oscillation. Panel C2 shows multiple simulations of the macroscale particle system with the value of $\bar{J}$ varying between simulations and $\beta = 0.4$ fixed.

A.5. Figure 5. Figure 5 presents simulations of the mesoscale finite-size model and the corresponding GKEs on the periodic spatial domain $\Gamma = \mathcal{S}_5$ (represented by the interval $[0, 5]$ with the endpoints identified).

For panels A1, A2, and A3 we directly simulated the finite-size system using the Gillespie algorithm. Panels A4, A5, and A6 are solutions of the GKEs of the mean-field limit. In particular, we solve the nonlinear integro-differential equation

$$\begin{aligned}(\text{A.1}) \quad \frac{\partial}{\partial t} G(t, r) &= (1 - G(t, r)) \phi \left(\int_{\Gamma} W(r', r) G(t, r') dr' \right) \\ &\quad - G(t, r) \left(1 - \int_{\Gamma} J(r', r) G(t, r') dr' \right), \quad (t, r) \in \mathbb{R}_+ \times \Gamma,\end{aligned}$$

with periodic boundary conditions and kernels as described in section 3.2.1. We solve equation (A.1) by discretizing time with an explicit Euler scheme to obtain

$$(A.2) \quad G(n+1, r) = G(n, r) + h \left\{ (1 - G(n, r)) \phi \left(\int_{\Gamma} W(r', r) G(n, r') dr' \right) - G(n, r) \left(1 - \int_{\Gamma} J(r', r) G(n, r') dr' \right) \right\}, \quad (n, r) \in \mathbb{Z}^+ \times \Gamma.$$

We then discretize the integrals using the 1D trapezoidal rule and approximate the solution on the evenly spaced grid $\{0, \Delta, 2\Delta, \dots, 5\}$ for some $\Delta > 0$. In practice, we found that a step size $h < 0.1$ and 200 spatial grid points was sufficient to ensure numerical stability of the scheme, and the scheme remained stable as we decreased the time step and increased the number of grid points.

In panels B1–B3 of Figure 5 we compare the results of simulations of the mesoscale particle system and the corresponding generalized Kolmogorov equations with a *nonuniform initial site distribution*. Panel B1 is a space-time plot of the result from a direct simulation of the mesoscale particle system with $N = 2000$ sites whose initial positions are drawn according to the distribution

$$(A.3) \quad dq(x) = (a + bx) \mathbb{1}_{[0, L]}(x) dx, \quad a, b > 0, \quad x \in [0, 1],$$

with $a = 0.4$ and $b = 1.2$ for this particular example. The spatial domain Γ is now the interval $[0, 1]$, and the dispersal kernels are standard Gaussians and of convolution type, i.e.,

$$W(r, r') = \frac{J(r, r')}{\bar{J}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(r-r')^2}{2\sigma^2}}, \quad r, r' \in \mathbb{R}.$$

We take $\bar{J} = 1.1$ in these simulations. In contrast to panel A, which has periodic boundary conditions, the boundary conditions for this simulation are “reflecting” in the following sense: consider a function u defined on $[0, L]$. First, extend u from a function on $[0, L]$ to $[-L, L]$ by reflection:

$$u_R(x, t) = \begin{cases} u(x, t), & x \in [0, L], \\ u(-x, t), & x \in [-L, 0]. \end{cases}$$

Now let $\tilde{u}$ be the standard $2L$ periodic extension of u_R (as defined above) so that $\tilde{u}$ is defined on all of $\mathbb{R}$. When the system has a heterogeneous spatial structure as in this example, since sites are less densely packed close to zero and more densely packed situated closer to L , a periodic extension introduces unrealistic boundary effects. In particular, periodic boundaries put dense regions adjacent to sparse site regions, while the reflecting boundary makes sure that sparse regions are neighbors of sparse regions at the boundaries. When dispersal kernels are sufficiently local, this avoids unintuitive and physically unrealistic solutions with significant boundary effects. For the simulation of the particle system, we once more use the Gillespie algorithm to simulate transition times.

Panel B2 of Figure 5 is a space-time plot of the solution of the GKEs for the mean-field mesoscale system with initial site distribution given by (A.3). The integro-differential

equations to be solved are given by

$$(A.4) \quad \frac{\partial}{\partial t} G(t, r) = (1 - G(t, r)) \phi \left(\int_{\Gamma} (a + b r') W(r - r') G(t, r') dr' \right) \\ - G(t, r) \left(1 - \int_{\Gamma} (a + b r') J(r - r') G(t, r') dr' \right), \quad (t, r) \in \mathbb{R}_+ \times \Gamma,$$

with reflecting boundary conditions as outlined above. We solved (A.4) numerically using the same Euler time discretization and trapezoidal rule spatial discretization as before with similar discretization parameters giving good numerical stability of the scheme.

Panel B3 has three components overlaid on the same axes: the time $t = 500$ state of the particle system from B1 (solid blue), the time $t = 500$ state of the GKE solution from B2 (dashed black line), and the codimension-1 bifurcation diagram in r of the ODE

$$(A.5) \quad \frac{d}{dt} G(t) = (1 - G(t)) (a + b r) G(t) - G(t) (1 - (a + b r) \bar{J} G(t)).$$

The ODE given by (A.5) is obtained from the GKE (A.4) by formally letting $\sigma \rightarrow 0^+$ in both the fire and dispersal kernels in order to study the case of no spatial interaction between sites. Hence (A.5) can be considered as the natural zero dispersal limit of the system with Gaussian kernels. The bifurcation diagram for (A.5) is overlaid with solid red lines for stable equilibria, and unstable equilibria are denoted by dashed red lines.

Acknowledgment. The authors sincerely thank the anonymous reviewers for their constructive remarks which have considerably improved the presentation and clarity of the paper.

REFERENCES

- [1] R. AGHAJANI, P. ROBERT, AND W. SUN, *A large scale analysis of unreliable stochastic networks*, Ann. Appl. Probab., 28 (2018), pp. 851–887.
- [2] N. ANDELA, D. MORTON, L. GIGLIO, Y. CHEN, G. VAN DER WERF, P. KASIBHATLA, R. DEFRIES ET AL., *A human-driven decline in global burned area*, Science, 356 (2017), pp. 1356–1362.
- [3] D. BERNOULLI, *Hydrodynamica: Sive De Viribus et Motibus Fluidorum Commentarii*, Johannis Reinholdi Dulseckeri, 1738.
- [4] P. BILLINGSLEY, *Convergence of Probability Measures*, John Wiley & Sons, 2013.
- [5] L. BOLTZMANN, *Über das Wärmegleichgewicht zwischen mehratomigen Gasmolekülen*, k.k. Hof- und Staatsdruckerei, 1871.
- [6] F. BORGOGNO, P. D'ODORICO, F. LAIO, AND L. RIDOLFI, *Mathematical models of vegetation pattern formation in ecohydrology*, Rev. Geophys., 47 (2009), RG1005.
- [7] T. CABANA AND J. D. TOUBOUL, *Large deviations for randomly connected neural networks I: Spatially extended systems*, Adv. Appl. Probab., 50 (2018), pp. 944–982.
- [8] V. CAPASSO, *Mathematical Structures of Epidemic Systems*, Lecture Notes in Biomath. 97, Springer-Verlag, 1993.
- [9] J. S. CLARK, M. SILMAN, R. KERN, E. MACKLIN, AND J. HILLERISLAMBERS, *Seed dispersal near and far: Patterns across temperate and tropical forests*, Ecology, 80 (1999), pp. 1475–1494.
- [10] R. CLAUSIUS, *Über die Art der Bewegung, welche wir Wärme nennen*, Ann. Phys., 176 (1857), pp. 353–380.
- [11] J. COX AND R. DURRETT, *Nonlinear voter models*, in Random Walks, Brownian Motion, and Interacting Particle Systems, Springer, 1991, pp. 189–201.

- [12] J. T. COX, R. DURRETT, AND E. A. PERKINS, *Voter model perturbations and reaction diffusion equations*, *Astérisque*, no. 349 (2013).
- [13] P. DAI PRA AND F. DEN HOLLANDER, *McKean–Vlasov limit for interacting random processes in random media*, *J. Statist. Phys.*, 84 (1996), pp. 735–772.
- [14] J. N. DARROCH AND E. SENETA, *On quasi-stationary distributions in absorbing continuous-time finite Markov chains*, *J. Appl. Probab.*, 4 (1967), pp. 192–196.
- [15] D. A. DAWSON, J. TANG, AND Y. Q. ZHAO, *Balancing queues by mean field interaction*, *Queueing Syst.*, 49 (2005), pp. 335–361.
- [16] D. A. DAWSON AND X. ZHENG, *Law of large numbers and central limit theorem for unbounded jump mean-field models*, *Adv. Appl. Math.*, 12 (1991), pp. 293–326.
- [17] R. L. DOBRUSHIN, *Prescribing a system of random variables by conditional distributions*, *Theory Probab. Appl.*, 15 (1970), pp. 458–486, <https://doi.org/10.1137/1115049>.
- [18] R. DURRETT, *Coexistence in stochastic spatial models*, *Ann Appl. Probab.*, 19 (2009), pp. 477–496.
- [19] R. DURRETT AND S. A. LEVIN, *Spatial aspects of interspecific competition*, *Theor. Popul. Biol.*, 53 (1998), pp. 30–43.
- [20] R. DURRETT AND R. MA, *A Heterogeneous Spatial Model in which Savanna and Forest Coexist in a Stable Equilibrium*, preprint, <https://arxiv.org/abs/1808.08159>, 2018.
- [21] R. DURRETT AND C. NEUHAUSER, *Particle systems and reaction-diffusion equations*, *Ann. Probab.*, (1994), pp. 289–333.
- [22] R. DURRETT AND Y. ZHANG, *Coexistence of grass, saplings and trees in the Stayer–Levin forest model*, *Ann. Appl. Probab.*, 25 (2015), pp. 3434–3464.
- [23] B. ERMENTROUT, *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*, *Software Environments Tools 14*, SIAM, 2002, <https://doi.org/10.1137/1.9780898718195>.
- [24] S. FENG AND X. ZHENG, *Solutions of a class of nonlinear master equations*, *Stochastic Process. Appl.*, 43 (1992), pp. 65–84.
- [25] R. A. FISHER, C. D. KOVEN, W. R. ANDEREGG, B. O. CHRISTOFFERSEN, M. C. DIETZE, C. E. FARRIOR, J. A. HOLM ET AL., *Vegetation demographics in Earth system models: A review of progress and priorities*, *Global Change Biol.*, 24 (2018), pp. 35–54.
- [26] E. GILAD, J. VON HARDENBERG, A. PROVENZALE, M. SHACHAK, AND E. MERON, *Ecosystem engineers: From pattern formation to habitat creation*, *Phys. Rev. Lett.*, 93 (2004), 098105.
- [27] N. GOEL, V. GUTTAL, S. A. LEVIN, AND A. C. STAYER, *Dispersal increases the resilience of tropical savanna and forest distributions*, *The American Naturalist*, 195 (2020), pp. 833–850.
- [28] K. GOWDA, H. RIECKE, AND M. SILBER, *Transitions between patterned states in vegetation models for semiarid ecosystems*, *Phys. Rev. E*, 89 (2014), 022701.
- [29] C. GRAHAM AND P. ROBERT, *Interacting multi-class transmissions in large stochastic networks*, *Ann. Appl. Probab.*, 19 (2009), pp. 2334–2361.
- [30] A. HASTINGS, K. C. ABBOTT, K. CUDDINGTON, T. FRANCIS, G. GELLNER, Y.-C. LAI, A. MOROZOV, S. PETROVSKII, K. SCRANTON, AND M. L. ZEEMAN, *Transient phenomena in ecology*, *Science*, 361 (2018), p. eaat6412.
- [31] N. IKEDA AND S. WATANABE, *Stochastic Differential Equations and Diffusion Processes*, 2nd ed., North-Holland Math. Library 24, North-Holland, 1992.
- [32] T. H. KEITT, M. A. LEWIS, AND R. D. HOLT, *Allee effects, invasion pinning, and species borders*, *The American Naturalist*, 157 (2001), pp. 203–216.
- [33] C. A. KLAUSMEIER, *Regular and irregular patterns in semiarid vegetation*, *Science*, 284 (1999), pp. 1826–1828.
- [34] R. LEFEVER AND O. LEJEUNE, *On the origin of tiger bush*, *Bull. Math. Biol.*, 59 (1997), pp. 263–294.
- [35] Q. LI, A. C. STAYER, W. E, AND S. A. LEVIN, *Spatial feedbacks and the dynamics of savanna and forest*, *Theor. Ecol.*, 12 (2019), pp. 237–262.
- [36] A. L. LLOYD AND R. M. MAY, *Spatial heterogeneity in epidemic models*, *J. Theoret. Biol.*, 179 (1996), pp. 1–11.
- [37] E. LUÇON AND W. STANNAT, *Mean field limit for disordered diffusions with singular interactions*, *Ann. Appl. Probab.*, 24 (2014), pp. 1946–1993.
- [38] H. P. MCKEAN, JR., *A class of Markov processes associated with nonlinear parabolic equations*, *Proc.*

- Natl. Acad. Sci. USA, 56 (1966), pp. 1907–1911.
- [39] J. D. MURRAY, *Mathematical Biology*, Springer, New York, 2004.
 - [40] C. NEUHAUSER, *A long range sexual reproduction process*, Stochastic Process. Appl., 53 (1994), pp. 193–220.
 - [41] J. R. NORRIS, *Markov Chains*, reprint of 1997 original, Cambridge Ser. Statist. Probab. Math. 2, Cambridge University Press, 1998.
 - [42] C. QUIÑINAO, A. PROCHIANTZ, AND J. TOUBOUL, *Local homeoprotein diffusion can stabilize boundaries generated by graded positional cues*, Development, 142 (2015), pp. 1860–1868.
 - [43] D. REVUZ AND M. YOR, *Continuous Martingales and Brownian Motion*, Grundlehren Math. Wiss. 293, Springer-Verlag, 1999.
 - [44] P. ROBERT AND J. D. TOUBOUL, *On the dynamics of random neuronal networks*, J. Statist. Phys., 165 (2016), pp. 545–584.
 - [45] E. SCHERTZER, A. C. STAYER, AND S. A. LEVIN, *Implications of the spatial dynamics of fire spread for the bistability of savanna and forest*, J. Math. Biol., 70 (2015), pp. 329–341.
 - [46] M. SCHEUTZOW, *Periodic behavior of the stochastic Brusselator in the mean-field limit*, Probab. Theory Relat. Fields, 72 (1986), pp. 425–462.
 - [47] A. C. STAYER, S. ARCHIBALD, AND S. A. LEVIN, *Tree cover in sub-Saharan Africa: Rainfall and fire constrain forest and savanna as alternative stable states*, Ecology, 92 (2011), pp. 1063–1072.
 - [48] A. C. STAYER, G. P. ASNER, I. RODRIGUEZ-ITURBE, S. A. LEVIN, AND I. P. SMIT, *Spatial patterning among savanna trees in high-resolution, spatially extensive data*, Proc. Natl. Acad. Sci. USA, 116 (2019), pp. 10681–10685.
 - [49] A. C. STAYER AND S. LEVIN, *Integrating theoretical climate and fire effects on savanna and forest systems*, The American Naturalist, 180 (2012), pp. 211–224.
 - [50] A.-S. SZNITMAN, *Topics in propagation of chaos*, in Ecole d'Été de Probabilités de Saint-Flour XIX–1989, Springer, 1991, pp. 165–251.
 - [51] M.-N. THAI, *Birth and Death Process in Mean Field Type Interaction*, preprint, <https://arxiv.org/abs/1510.03238>, 2015.
 - [52] S. THOMPSON AND G. KATUL, *Plant propagation fronts and wind dispersal: An analytical model to upscale from seconds to decades using superstatistics*, The American Naturalist, 171 (2008), pp. 468–479.
 - [53] J. D. TOUBOUL, *Propagation of chaos in neural fields*, Ann. Appl. Probab., 24 (2014), pp. 1298–1328.
 - [54] J. D. TOUBOUL, *Spatially extended networks with singular multi-scale connectivity patterns*, J. Statist. Phys., 156 (2014), pp. 546–573.
 - [55] J. TOUBOUL, G. HERMANN, AND O. FAUGERAS, *Noise-induced behaviors in neural mean field dynamics*, SIAM J. Appl. Dyn. Syst., 11 (2012), pp. 49–81, <https://doi.org/10.1137/110832392>.
 - [56] J. D. TOUBOUL, A. C. STAYER, AND S. A. LEVIN, *On the complex dynamics of savanna landscapes*, Proc. Natl. Acad. Sci. USA, 155 (2018), pp. E1336–E1345.
 - [57] I. A. VAN DE LEEMPUT, E. H. VAN NES, AND M. SCHEFFER, *Resilience of alternative states in spatially extended ecosystems*, PLoS ONE, 10 (2015), e0116859.
 - [58] C. VILLANI, *A review of mathematical topics in collisional kinetic theory*, in Handbook of Mathematical Fluid Dynamics, Vol. I, North-Holland, 2002, pp. 71–305.
 - [59] B. WUYTS, A. R. CHAMPNEYS, N. VERSCHUEREN, AND J. I. HOUSE, *Tropical tree cover in a heterogeneous environment: A reaction-diffusion model*, PLoS ONE, 14 (2019), pp. 1–16.