Mathematical Models and Methods in Applied Sciences
Vol. 13, No. 4 (2003) 491–525

APPROXIMATE REDUCTION OF MULTI-TYPE GALTON–WATSON PROCESSES WITH TWO TIME SCALES

LUIS SANZ

Departamento de Matemáticas, E.T.S.I. Industriales,
Universidad Politécnica de Madrid, José Gutiérrez Abascal, 2, 28006 Madrid, Spain
lsanz@etsii.upm.es

ÁNGEL BLASCO* and RAFAEL BRAVO DE LA PARRA†

Departamento de Matemáticas, Universidad de Alcalá,
28871 Alcalá de Henares (Madrid), Spain
**angel.blasco@uah.es*
†*rafael.bravo@uah.es*

Received 24 April 2002
Communicated by P. Auger

Approximate aggregation techniques consist of introducing certain approximations that allow one to reduce a complex system involving many coupled variables obtaining a simpler "aggregated system" governed by a few "macrovariables". Moreover, they give results that allow one to extract information about the complex original system in terms of the behavior of the reduced one. Often, the feature that allows one to carry out such a reduction is the presence of different time scales.

In this work we deal with the approximate aggregation of a model for a population subjected to demographic stochasticity and whose dynamics is controlled by two processes with different time scales. There are no restrictions on the slow process while the fast process is supposed to be conservative of the total number of individuals. The incorporation of the effects of demographic stochasticity in the dynamics of the population makes both the fast and the slow processes being modelled by two multi-type Galton–Watson branching processes. We present a multi-type global model that incorporates the combined effect of the fast and slow processes and develop a method that takes advantage of the difference of time scales to reduce the model obtaining an "aggregated" simpler system. We show that, given the separation of time scales between the two processes is high enough, we can obtain relevant information about the behavior of the multi-type global model through the study of this simple aggregated system.

Keywords: Approximate aggregation; multi-type Galton–Watson processes; time scales.

AMS Subject Classification: 92D25, 37N25

1. Introduction

Approximate aggregation techniques provide a methodology for studying the dynamics of some high dimensional dynamical systems by means of a lower dimensional one. We deal with systems that are complex in the sense of having a large number of variables, and make use of the existence of different time scales (i.e. of biological processes which take place with characteristic times very different from each other) to introduce approximations that allow substitution of the global system by a reduced system that retains at least some of the properties of the original system.

We can think of a system with different time scales as an hierarchically structured system with division into subsystems that are weakly coupled and simultaneously exhibit a strong internal dynamics. The idea of aggregation is to choose a few (usually one) "macrovariables" for each subsystem and to build a reduced system for these macrovariables. In many cases the dimension of the aggregated system is considerably lower than that of the original system.

These techniques have been widely studied in many different contexts: discrete and continuous time models based on linear and nonlinear systems in both autonomous and non-autonomous cases (see for example Refs. 1, 2, 6, 12, 17 and 19 where further references on the field can be found).

Currently, stochasticity is one of the main items in the field of ecological modelling. In general, ecosystems are influenced by a large number of factors and this makes it impractical to include all of them in a mathematical model. On the other hand, its exclusion generates unexplained variation whose consideration implies the use of stochasticity. In population models, this stochasticity appears in two ways:

(a) When it affects the vital rates of the population and is supposed to arise from random changes in the environment (climatic conditions, interaction with other species, etc.) it is called "environmental stochasticity".[20] The study of aggregation techniques for models subjected to environmental stochasticity has been addressed by Sanz and Bravo de la Parra.[18]
(b) When it produces deviations in the behavior of each individual with respect to the global vital rates, it is called demographic stochasticity. This second kind of stochasticity becomes essential in the study of small populations, where its effects may be crucial for the fate of the community, sometimes even causing its total extinction.[10]

The classical mathematical tools for modelling the dynamics of a structured population subjected to demographic stochasticity in discrete time are the multi-type Galton–Watson branching processes.[13,16] These models have been frequently used in genetics, physics, mathematics and biology. Some recent contributions in the field can be found in Refs. 5, 7, 15 and 21. The effects of environmental stochasticity can be incorporated in these models, obtaining the so-called Multi-type Branching Processes in Random Environments (MBPRE).[3,8]

Aggregation techniques in the context of systems subjected to demographic stochasticity have been introduced in Ref. 4 to deal with a population without age structure and living in a multipatch environment. The resulting multi-type g.w.p. is reduced to obtain a unitype process for the total size of the population. Moreover, results are given that allow one to obtain information of the global process in terms of the reduced one. Specifically, both the mean population vector and the probability of extinction of the original model can be approximated by those corresponding to the reduced unitype model.

This work extends the aggregation techniques presented in Ref. 4 in two directions: in the first place, we contemplate the approximate aggregation of a general g.w.p. in which there are two time scales involved, being the resulting reduced system a simpler multi-type g.w.p. governed by a set of so called "macrovariables". We present a "global" model for a system whose dynamics is controlled by two processes with different time scales. Essentially, the slow process can correspond to any general multi-type g.w.p. meanwhile the fast dynamics is associated to a process which is conservative of the total number of individuals and is modelled by a Markov chain. By making use of the existence of different time scales we introduce approximations that allow us to reduce the global model. As an illustration of the use of our aggregation technique, we undertake the reduction of a multiregional age-structured model subjected to demographic stochasticity.

In the second place, we extend the analysis of the relationships between the original system and the reduced model carried out in Ref. 4. We obtain relationships between the two systems in our general setting in the following sense: if the separation of time scales between the fast and the slow processes is large enough, then certain features of the original system can be approximated by those corresponding to the reduced one. These features include the mean and the moments of second order of the population vector, the probability of extinction and, for the subcritical case, the mean population vector conditional on non-extinction. As a consequence of these results it follows that we can approximate the behavior of the original complex model by studying the reduced system. Therefore, the aggregation technique we propose constitutes a tool for the analysis of complex models with two time scales and subjected to the effects of demographic stochasticity.

2. Organization of the Work and Main Results

Section 3 is devoted to motivate the aggregation procedure developed in the paper. In order to do so, we set out a simple multiregional model which stands out as an important particular case of the general model introduced below. It describes the dynamics of an age-structured population with two age classes living in a two-patch environment and subjected to the effects of demographic stochasticity.

In Sec. 4, we present a general multi-type g.w.p. for a structured population whose dynamics is controlled by two processes with different time scales, to which we will refer as fast and slow dynamics. We assume these processes to be subjected to demographic stochasticity and, consequently, they are modelled by multi-type g.w.p.'s. The model we propose can be considered as a generalization, that takes into account the effects of demographic stochasticity, of a model previously presented by the authors[17] corresponding to a system with two time scales in a deterministic context.

The population is supposed to be structured in q groups attending to any characteristic of the individuals and each group is divided into subgroups in such a way that the total number of subgroups is N. We impose no restrictions over the slow process, meanwhile for the fast process we assume: (a) is internal for the groups, i.e. there is no transference of individuals between the different groups by means of the fast process and (b) is conservative of the total number of individuals and consequently it is modelled through a Markov chain.

We introduce these two g.w.p.'s through their probability generating functions (in the sequel p.g.f.'s) each of which is referred to the projection interval of the corresponding process. Then we formulate a multi-type g.w.p. with N types that takes into account the effect of both dynamics and is referred to the characteristic time of the slow process.

In order to approximate the effect of the fast process over the time step of the model, which is much longer than its own projection interval, we assume that in each time step of the model the fast process acts in a large number k of times. Here k can be interpreted as the ratio of the characteristic times for the slow and the fast processes. In this way, the model that takes into account the joint effect of the fast and the slow process referred to the characteristic time of the latter can be interpreted as the "composition" of k iterations of the fast process followed by one iteration of the slow process. Lemma 1 characterizes mathematically the composition of multi-type g.w.p.'s and allows us to build the p.g.f. and the matrices that characterize the first and second order moments for the population vector of the global model.

In Sec. 5 we carry out the reduction of the global model in two steps. In the first place we approximate the original system by an auxiliary system in which the Markov chain modelling the fast process reaches its equilibrium distribution in each time step of the model. By defining the macrovariables as the total population in each group, we can reduce this auxiliary system obtaining a simpler aggregated g.w.p. with q types. Propositions 1 and 2 allow one to construct this reduced system in terms of the slow process and the equilibrium characteristics of the fast process. Indeed, Proposition 1 gives the p.g.f., whose components are weighted sums of the p.g.f.'s corresponding to the slow process being the weights dependent on the equilibrium probabilities of the fast dynamics. On the other hand, Proposition 2 provides the matrices characterizing the moments of first and second order for the reduced system. The last part of this section deals with the application of the aggregation procedure just described to the multiregional model introduced in Sec. 3.

The behaviors of the original model and the reduced model are compared in Sec. 6. Proposition 3, which is a direct application of some previous results by the authors[17] regarding aggregation in a deterministic context, states that the expected growth rate of the population size for the global process can be obtained as a perturbation of the expected growth rate for the reduced model. Moreover, the dominant eigenvectors of the matrix of expected values for the global system can be approximated in terms of those corresponding to the aggregated system. These perturbations are characterized asymptotically when k, that measures the separation of time scales between the slow process and the fast process, tends to infinity. The size of this perturbation is related to the modulus of the subdominant eigenvalue of the matrix of expected values for the fast process, in such a way that the faster the fast process reaches its equilibrium distribution, the smaller the perturbation is.

Proposition 4 relates the properties of singularity, positive regularity and super/subcriticality of the g.w.p.'s corresponding to the global system and the reduced model. In particular, the latter is supercritical (subcritical) if and only if the latter is for a high enough separation of time scales between the fast and the slow process.

In Proposition 5 we show that the probability of extinction in finite time in the global process can be obtained as a perturbation of the corresponding probability of extinction in the reduced system, while Proposition 6, which is a generalization of a similar result of the authors in Ref. 4, relates the probabilities of ultimate extinction for both systems.

Proposition 7 allows us to approximate the asymptotic behavior of the mean population vector for the global model in terms of the dominant eigenvalues and eigenvectors of the matrix of expected values for the reduced model, meanwhile Proposition 8 approximates the asymptotic behavior of the second order moments of the original system in terms of information pertaining to the aggregated system.

As guaranteed by Proposition 4, in the case the aggregated system is subcritical then so is the original model for a sufficiently high separation of time scales, and consequently in both models population becomes eventually extinct. However in both cases the expected population vector conditional on non-extinction approaches a constant vector. Proposition 9, which constitutes our main result as far as the relationships between the original and the reduced systems is concerned, allows one to approximate the asymptotic behavior of the expected population vector conditional on non-extinction in terms of the characteristics of the aggregated system.

In order to estimate the accuracy of the approximations mentioned above, i.e. of the error we incur when studying the global system in terms of the reduced one, we have carried out some computer simulations for the multiregional model of Sec. 3. These show that the error is small even for moderate values of k and so our reduction technique can be useful in the study of real biological populations.

3. A Multiregional Model with Fast Migration

In this section we consider a multiregional model for an age-structured population with two age classes, young and adults. The population is distributed between two spatial patches and influenced by the effects of demographic stochasticity. Within each age class (group) we classify the individuals into subgroups regarding the patch (T1 or T2) they live in. For each n, let x_n^{ij} be the number of individuals of age i living in patch j at time n. Then the population vector at time n is $\mathbf{X}_n = (x_n^{11}, x_n^{12}, x_n^{21}, x_n^{22})^{\mathrm{T}}$ and so the random sequence $\mathbf{X}_n$ is a realization of a multi-type Galton–Watson branching process with four types.

The evolution of the population is governed by two processes: a birth–death process corresponding to demography that models the transition among the two age classes, and a Markov chain that models the migration among the two patches. We assume that, as is often the case in real populations[1,17] migration is a fast process with respect to demography.

We impose two hypotheses on the birth–death process modelling demography:

(HE1) The newborn offspring is supposed to inhabit the patch of the father.

(HE2) Each individual can produce at most two offspring in a time step.

These hypotheses stand for this particular model but they will be removed from the general setting described in Sec. 4.

Given a vector $\alpha = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}) \in \mathbb{Z}_+^4$ let $p_S^{ij}(\alpha)$ denote the probability of an individual of age i living at patch j to produce an offspring α (i.e. α_{11} young at T1, α_{12} at T2, α_{21} adults at T1 and α_{22} at T2) after an iteration of the birth–death process.

Note that hypothesis (HE1) implies that $p_S^{i1}(\alpha) = 0$ if $\alpha_{12} > 0$ or $\alpha_{22} > 0$ and that $p_S^{i2}(\alpha) = 0$ if $\alpha_{11} > 0$ or $\alpha_{21} > 0$. Besides, from hypothesis (HE2), $p_S^{i2}(\alpha) = 0$ if $\alpha_{11} > 2$ or $\alpha_{12} > 0$. On the other hand, a young individual may produce an adult by surviving to the next generation but it will never produce more than one, so $p_S^{1j}(\alpha) = 0$ if $\alpha_{2j} > 1$. Finally, an adult can just produce young individuals, so $p_S^{2j}(\alpha) = 0$ if $\alpha_{2j} > 0$. From these remarks we deduce that the p.g.f. associated to the birth–death process in our example has the form:

$$G_S^{11}(\mathbf{s}) = p_{0000}^{11} + p_{1000}^{11} s_{11} + p_{2000}^{11} s_{11}^2 + p_{0010}^{11} s_{21} + p_{1010}^{11} s_{11} s_{21} + p_{2010}^{11} s_{11}^2 s_{21} \,,$$

$$G_S^{12}(\mathbf{s}) = p_{0000}^{12} + p_{0100}^{12} s_{12} + p_{0200}^{12} s_{12}^2 + p_{0001}^{12} s_{22} + p_{0101}^{12} s_{12} s_{22} + p_{0201}^{12} s_{12}^2 s_{22} \,,$$

$$G_S^{21}(\mathbf{s}) = p_{0000}^{21} + p_{1000}^{21} s_{11} + p_{2000}^{21} s_{11}^2 \,; \quad G_S^{22}(\mathbf{s}) = p_{0000}^{22} + p_{0100}^{22} s_{12} + p_{0200}^{22} s_{12}^2 \,,$$

defined for $\mathbf{s} = (s_{11}, s_{12}, s_{21}, s_{22}) \in \mathbb{R}^4$ where, in order to keep the notation compact, we are denoting $p_{1010}^{11} := p_S^{11}(1, 0, 1, 0)$ and so on.

Besides, the matrix of expected values associated to the birth–death process is $\mathbf{M} = (m_{ij}) \in \mathbb{R}^{4\times 4}$ where

$$m_{11} = p_{1000}^{11} + 2p_{2000}^{11} + p_{1010}^{11} + 2p_{2010}^{11} \,; \quad m_{31} = p_{0010}^{11} + p_{1010}^{11} + p_{2010}^{11} \,,$$

$$m_{22} = p_{0100}^{12} + 2p_{0200}^{12} + p_{0101}^{12} + 2p_{0201}^{12} \,; \quad m_{42} = p_{0001}^{12} + p_{0101}^{12} + p_{0201}^{12} \,,$$

$$m_{13} = p_{1000}^{21} + 2p_{2000}^{21} \,; \quad m_{24} = p_{0100}^{22} + 2p_{0200}^{22}$$

and the remaining components are null.

Regarding migration, we assume that this is a conservative process of the total number of individuals in each of the age classes. Therefore, in each of these age classes it is modelled by a Markov chain with two states corresponding to the two patches. For each vector $\alpha = (\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}) \in \mathbb{Z}_+^4$, let $p_F^{ij}(\alpha)$ denote the probability of an i-aged individual living in patch j to produce an offspring α after an iteration of the migration process. The p.g.f. associated to migration is

$$G_F^{11}(\mathbf{s}) = p_F^{11}(\mathbf{e}^{11})s_{11} + p_F^{11}(\mathbf{e}^{12})s_{12}\,; \quad G_F^{12}(\mathbf{s}) = p_F^{12}(\mathbf{e}^{11})s_{11} + p_F^{12}(\mathbf{e}^{12})s_{12}\,,$$

$$G_F^{21}(\mathbf{s}) = p_F^{21}(\mathbf{e}^{21})s_{21} + p_F^{21}(\mathbf{e}^{22})s_{22}\,; \quad G_F^{22}(\mathbf{s}) = p_F^{22}(\mathbf{e}^{21})s_{21} + p_F^{22}(\mathbf{e}^{22})s_{22}\,,$$

and is defined for $\mathbf{s} \in \mathbb{R}^4$, where $\mathbf{e}^{ij}$ is the vector that describes a population composed of a single individual with age i and living in patch j. Then, the matrices of expected values for migration in each age class are

$$\mathbf{P}_i = \begin{pmatrix} p_F^{i1}(\mathbf{e}^{i1}) & p_F^{i2}(\mathbf{e}^{i1}) \\ p_F^{i1}(\mathbf{e}^{i2}) & p_F^{i2}(\mathbf{e}^{i2}) \end{pmatrix}\,; \quad i = 1, 2$$

and for the whole population we have matrix $\mathbf{P} = \mathrm{diag}\{\mathbf{P}_1, \mathbf{P}_2\}$. We assume that $\mathbf{P}_1$ and $\mathbf{P}_2$ are primitive.

Now, we formulate a global model that takes into account the joint effect of both demography and migration in the following way: we set the time step of the model as the projection interval of the birth–death process, and we assume that, in each of these time steps, migration acts k times before the birth–death process does. Here k can be interpreted as the ratio of the characteristic times of demography and migration. Then, according to Lemma 1 bellow, the p.g.f. of the original system is given by $\mathbf{G}_k(\mathbf{s}) = \mathbf{G}_F \circ \overset{(k)}{\cdots} \circ \mathbf{G}_F \circ \mathbf{G}_S$, and the matrix of expectations for this global model is $\mathbf{MP}^k$.

In this work we propose a reduction technique that allows us to study this four-dimensional model by means of a two-dimensional one, called "aggregated model". In order to do so, let us first outline the former problem in a more general context. In Sec. 5, we will describe the general aggregation procedure and apply it to this particular case.

4. A Multi-Type Galton–Watson Process with Two Time Scales

We consider a stage-structured population in which the population is classified into stages or groups attending to any characteristic of the life cycle. Moreover, each of these groups is divided into several subgroups that may correspond to different spatial patches, different individual activities or any other characteristic that could change the life cycle parameters. The model is general in the sense that we do not state in detail the nature of the groups or the subgroups. Let q be the number of groups and assume that each group $i = 1, 2, \ldots, q$ is itself split into N_i subgroups. Therefore, the total number of subgroups is $N = N_1 + N_2 + \cdots + N_q$.

We denote by x_n^{ij} the number of individuals in state (i, j), i.e. the number of individuals in subgroup j of group i at time n; $i = 1, 2, \ldots, q$ and $j =$

$1, 2, \ldots, N_i$. The composition of the total population is then given by vector $\mathbf{X}_n = (x_n^{11}, \ldots, x_n^{1N_1}, \ldots, x_n^{q1}, \ldots, x_n^{qN_q})^{\mathrm{T}} \in \mathbb{R}^N$, where "T" denotes transposition.

We assume that the evolution of the population is governed by two processes whose corresponding characteristic time scales, and consequently their projection intervals are very different from each other.

Moreover, the system is supposed to be subjected to demographic stochasticity, i.e. the contribution of each individual at time n to the next generation takes place according to a given probability law for offspring production. For a given individual, this probability law is assumed to depend only on the state of the individual and not on the size of the population or on other individuals, and so the evolution of the system is modelled by means of a multi-type Galton–Watson branching process (in the sequel g.w.p.) $\mathbf{X}_0, \mathbf{X}_1, \ldots$.

In this way, the fast and the slow processes are modelled by two different g.w.p.'s, each of which is referred to the projection interval of the corresponding process. We choose as the projection interval of our model, that corresponding to the slow dynamics, i.e. the time elapsed between times n and $n+1$ is the projection interval of the slow process. For simplicity, we will denote the time span $[n, n+1)$ as Δ_n.

In the remaining of this section we introduce the fast and the slow processes through their probability generating functions (p.g.f.'s) and obtain the p.g.f. that characterizes the evolution of the population when both processes are operating.

4.1. *P.g.f. of the slow process*

In the following, $\alpha = (\alpha_{11}, \ldots, \alpha_{1N_1}, \ldots, \alpha_{q1}, \ldots, \alpha_{qN_q}) \in \mathbb{Z}_+^N$, i.e. α is a vector of $\mathbb{R}^N$ whose components are non-negative integers and $\mathbf{s} = (s_{11}, \ldots, s_{1N_1}, \ldots, s_{q1}, \ldots, s_{qN_q}) \in \mathbb{R}^N$. Throughout our work we will use the notation:

$$\mathbf{s}^\alpha := \prod_{ij} s_{ij}^{\alpha_{ij}} = s_{11}^{\alpha_{11}} \cdots s_{1N_1}^{\alpha_{1N_1}} \cdots s_{q1}^{\alpha_{q1}} \cdots s_{qN_q}^{\alpha_{qN_q}}\,.$$

Let $\mathbf{e}^{ij}$ be the canonical vector of $\mathbb{R}^N$ that describes a population composed of a single individual in state (i, j) and let us denote by $\| * \|_1$ the 1-norm in $\mathbb{R}^N$, i.e. if $\mathbf{z} = (z_1, z_2, \ldots, z_N)^{\mathrm{T}}$ we have $\|\mathbf{z}\|_1 = |z_1| + |z_2| + \cdots + |z_N|$.

Assume the system is controlled by the slow process exclusively, and let $\mathbf{h}_n = (h_n^{11}, \ldots, h_n^{1N_1}, \ldots, h_n^{q1}, \ldots, h_n^{qN_q})$ denote the population vector for this system at time n. Then we define

$$p_S^{ij}(\alpha) := pr(\mathbf{h}_{n+1} = \alpha \,|\, \mathbf{h}_n = \mathbf{e}^{ij})\,,$$

i.e. $p_S^{ij}(\alpha)$ is the probability of obtaining, by means of the slow process, α_{rs} individuals in state (r, s), $r = 1, \ldots, q; s = 1, \ldots, N_r$, from an individual in state (i, j). The only assumption we impose on the characteristics of the slow process is that, for each i and j, there is only a finite number of vectors α such that $p_S^{ij}(\alpha) > 0$.

Then, the p.g.f. for the slow process is the function $\mathbf{G}_S \colon \mathbb{R}^N \to \mathbb{R}^N$

$$\mathbf{G}_S(\mathbf{s}) = (G_S^{11}(\mathbf{s}), \ldots, G_S^{1N_1}(\mathbf{s}), G_S^{21}(\mathbf{s}), \ldots, G_S^{2N_2}(\mathbf{s}), \ldots, G_S^{q1}(\mathbf{s}), \ldots, G_S^{qN_q}(\mathbf{s}))$$

given by

$$G_S^{ij}(\mathbf{s}) = \sum_{\alpha} p_S^{ij}(\alpha)\mathbf{s}^{\alpha}\,; \quad i = 1, \ldots, q\,; \quad j = 1, \ldots, N_i$$

where the summation is extended to all $\alpha \in \mathbb{Z}_+^N$.

We denote as $\mathbf{M} \in \mathbb{R}^{N\times N}$, the matrix of expected values for the slow dynamics. We consider $\mathbf{M}$ divided into blocks $\mathbf{M}_{ij} = [M_{ij}^{rl}] \in \mathbb{R}^{N_i\times N_j}, 1 \leq i, j \leq q$, in such a way that $M_{ij}^{rl} = E(h_{n+1}^{ir} \,|\, \mathbf{h}_n = \mathbf{e}^{jl})$. Therefore,

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \cdots & \mathbf{M}_{1q} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_{q1} & \cdots & \mathbf{M}_{qq} \end{pmatrix},$$

where each block $\mathbf{M}_{ij}$ characterizes the expected values for the transference of individuals from group j to group i. In the same way, for each $u = 1, \ldots, q; v = 1, \ldots, N_u$, let

$$\begin{aligned} \mathbf{V}_S^{uv} &:= E(\mathbf{h}_{n+1}\mathbf{h}_{n+1}^{\mathrm{T}} \,|\, \mathbf{h}_n = \mathbf{e}^{uv}) \\ &\quad - E(\mathbf{h}_{n+1} \,|\, \mathbf{h}_n = \mathbf{e}^{uv})E(\mathbf{h}_{n+1}^{\mathrm{T}} \,|\, \mathbf{h}_n = \mathbf{e}^{uv}) \in \mathbb{R}^{N\times N} \end{aligned}$$

be the covariance matrix for the offspring produced by a parent in state (u, v), in the slow process. This is a symmetric matrix whose element in the position corresponding to states (i, r) and (j, l) is $(\mathbf{V}_S^{uv})_{ir,jl} = \mathrm{Cov}(h_{n+1}^{ir}, h_{n+1}^{jl} \,|\, h_n = e^{uv})$, i.e. the covariance between the offspring of states (i, r) and (j, l) given a parent in state (u, v).

4.2. *P.g.f. of the fast process*

Regarding the fast process we make the following assumptions:

(HA) The fast dynamics is an internal process for each group, i.e. there is no transference of individuals from one group to a different one.

(HB) Within each group $i = 1, \ldots, q$, the fast dynamics is a Markov chain with a primitive transition matrix $\mathbf{P}_i = [p_i^{rs}]$ (i.e. $\mathbf{P}_i^k$ is a positive matrix for some k) of dimensions $N_i \times N_i$. This fact can be expressed by saying that, within each group, the fast process corresponds to a positively regular singular multi-type branching process with matrix of expected values $\mathbf{P}_i$.

In this way, the matrix that characterizes the fast dynamics for the whole population is $\mathbf{P} = \mathrm{diag}(\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_q)$.

Assume for the moment that the system is controlled by the fast process exclusively, and let $\mathbf{w}_t$ denote the population vector for this system. Then, if the time span $[t, t+1)$ denotes the projection interval of the fast process, we have that for each $i = 1, \ldots, q; j = 1, \ldots, N_i$,

$$\begin{aligned} pr_F^{ij}(\alpha) &= pr(\mathbf{w}_{t+1} = \alpha \,|\, \mathbf{w}_t = \mathbf{e}^{ij}) \\ &= \begin{cases} p_i^{hj} & \text{if } \alpha = e^{ih} \text{ for some } h = 1, \ldots, N_i\,, \\ 0 & \text{otherwise}\,. \end{cases} \end{aligned} \tag{4.1}$$

The p.g.f. for the fast process is then a linear function $\mathbf{G}_F(\mathbf{s}) = (G_F^{11}(\mathbf{s}), \ldots, G_F^{1N_1}(\mathbf{s}), \ldots, G_F^{q1}(\mathbf{s}), \ldots, G_F^{qN_q}(\mathbf{s}))$ where, according to (4.1),

$$G_F^{ij}(\mathbf{s}) = \sum_{\alpha} p_F^{ij}(\alpha) s^{\alpha} = \sum_{h=1}^{N_i} p_i^{hj} s_{ih}\,, \quad i = 1, \ldots, q\,; \quad j = 1, \ldots, N_i\,.$$

Analogously, the probability for an individual initially in state (i, j) to be in state (i, s) after k periods corresponding to the fast process, is given by the corresponding entry of the matrix $\mathbf{P}_i^k = [(\mathbf{P}_i^k)_{rl}]$, i.e. $pr(\mathbf{w}_{t+k} = \mathbf{e}^{is} \,|\, \mathbf{w}_t = \mathbf{e}^{ij}) = (\mathbf{P}_i^k)_{sj}$, so that for all positive integers k,

$$pr(\mathbf{w}_{t+k} = \alpha \,|\, \mathbf{w}_t = \mathbf{e}^{ij}) = \begin{cases} (\mathbf{P}_i^k)_{sj} & \text{if } \alpha = e^{is} \text{ for some } s = 1, \ldots, N_i\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{4.2}$$

4.3. *P.g.f. of the global process*

As we stated before, we choose the projection interval of the slow process as the time step Δ_n of our "global" model. This model includes the effect of both the fast and the slow processes and, in order to formulate it, we need to approximate the effect of the former over a period much longer than its corresponding projection interval. In order to do so, we will assume that during each Δ_n, the fast process operates a number k of times before the slow process does, where k can be interpreted as the ratio between the projection intervals corresponding to the slow and fast dynamics. Since the projection intervals of both processes are supposed to be very different from each other, k is a large number and, moreover, we assume that it is an integer.

In order to determine the p.g.f. of this global model, we will study the "composition" of Galton–Watson processes. Let us consider a population with m types whose evolution is governed by two different g.w.p.'s $\mathbf{X}_n$ and $\mathbf{Y}_n$ in $\mathbb{R}^m$ characterized by p.g.f.'s $\mathbf{G}_{\mathbf{X}}$ and $\mathbf{G}_{\mathbf{Y}}$ respectively. We assume that, in each time step of the model, the population is first subjected to an iteration of the process with p.g.f. $\mathbf{G}_{\mathbf{X}}$ followed by an iteration of the process with p.g.f. $\mathbf{G}_{\mathbf{Y}}$. Therefore, we can think of the population as governed by a g.w.p. $\mathbf{Z}_n$ which can be interpreted as the "composition" of $\mathbf{X}_n$ and $\mathbf{Y}_n$. Then we have the following lemma, which generalizes a result for the unitype case in Ref. 13 to the multi-type case:

Lemma 1. (a) *The p.g.f.* $\mathbf{G}_Z$ *of the resulting process* Z *verifies* $\mathbf{G}_Z = \mathbf{G}_X \circ \mathbf{G}_Y$ ($\circ$ *denotes functional composition*).

(b) *Assume that the matrices of expected values for the processes* X *and* Y *are* $\mathbf{B} = [b_{ij}]$ *and* $\mathbf{T} = [t_{ij}]$ *respectively, i.e.* $b_{ij} = E(x_{n+1}^i \,|\, \mathbf{X}_n = e^j)$; $t_{ij} = E(y_{n+1}^i \,|\, \mathbf{Y}_n = e^j)$. *Then, the matrix* $\mathbf{A}$ *of expected values for the process* Z *verifies* $\mathbf{A} = \mathbf{TB}$.

(c) *For each* $j = 1, \ldots, m$, *let* $\mathbf{C}_X^j = [(C_X^j)_{rs}]$ *denote the matrix of second-order moments for the offspring production of a "parent" of type* j *in process* X, *i.e.* $(C_X^j)_{rs} = E(x_{n+1}^r x_{n+1}^s \,|\, \mathbf{X}_n = \mathbf{e}^j)$. *Let* $\mathbf{V}_Y^j$ *denote the covariance* *matrix of offspring production for a parent of type j in process Y, i.e.* $(V_Y^j)_{rs} = \mathrm{Cov}(y_{n+1}^r, y_{n+1}^s \,|\, \mathbf{Y}_n = \mathbf{e}^j)$. *Then the matrices* $\mathbf{C}_Z^j$ *and* $\mathbf{V}_Z^j$ *are given by*

$$\mathbf{C}_Z^j = \mathbf{T}\mathbf{C}_X^j\mathbf{T}^{\mathrm{T}} + \sum_{h=1}^m b_{hj}\mathbf{V}_Y^h\,, \quad \mathbf{V}_Z^j = \mathbf{C}_Z^j - \mathbf{T}\mathbf{B}\mathbf{e}^j(\mathbf{T}\mathbf{B}\mathbf{e}^j)^{\mathrm{T}}\,.$$

Proof. (a) and (b) are Lemma 2 in Ref. 4. The proof of (c) is obtained by differentiating twice in $\mathbf{G}_Z = \mathbf{G}_X \circ \mathbf{G}_Y$. Indeed, let j be fixed. In order to compute $\mathbf{C}_Z^j$ we need to evaluate the second derivatives of $\mathbf{G}_Z$

$$\frac{\partial^2 G_Z^j(\mathbf{s})}{\partial s_r \partial s_l} = \sum_{\alpha=1}^m \sum_{\beta=1}^m \frac{\partial^2 G_X^j(\mathbf{G}_Y(\mathbf{s}))}{\partial s_\alpha \partial s_\beta} \frac{\partial G_Y^\beta(\mathbf{s})}{\partial s_l} \frac{\partial G_Y^\alpha(\mathbf{s})}{\partial s_r} + \sum_{\alpha=1}^m \frac{\partial G_X^j(\mathbf{G}_Y(\mathbf{s}))}{\partial s_\alpha} \frac{\partial^2 G_Y^\alpha(\mathbf{s})}{\partial s_r \partial s_l}$$

in $\mathbf{s} = \mathbf{1}$. In the following we will use the notation $E[z_{n+1}^r \,|\, \mathbf{Z}_n = \mathbf{e}^j] \equiv E_j[z_{n+1}^r]$. We distinguish two cases: (i) First, let us consider the case where $r \neq l$. Then, taking into account that $\mathbf{G}_Y(\mathbf{1}) = \mathbf{1}$ and rearranging some terms, we obtain:

$$E_j[z_{n+1}^r z_{n+1}^l] = \sum_{\alpha=1}^m \sum_{\beta=1}^m E_j[x_{n+1}^\alpha x_{n+1}^\beta] E_\beta[y_{n+1}^l] E_\alpha[y_{n+1}^r] + \sum_{\alpha=1}^m E_j[x_{n+1}^\alpha](E_\alpha[y_{n+1}^r y_{n+1}^l] - E_\alpha[y_{n+1}^l] E_\alpha[y_{n+1}^r])$$

which can also be expressed as

$$(\mathbf{C}_Z^j)_{rl} = \sum_{\alpha=1}^m \sum_{\beta=1}^m t_{r\alpha} (\mathbf{C}_X^j)_{\alpha\beta} t_{l\beta} + \sum_{\alpha=1}^m b_{\alpha j} (\mathbf{V}_Y^\alpha)_{rl}\,.$$

(ii) Now let us consider the case $r = l$. Then

$$E_j[(z_{n+1}^r)^2] - E_j[z_{n+1}^r] = \sum_{\alpha=1}^m \sum_{\beta=1}^m E_j[x_{n+1}^\alpha x_{n+1}^\beta] E_\beta[y_{n+1}^l] E_\alpha[y_{n+1}^r] + \sum_{\alpha=1}^m E_j[x_{n+1}^\alpha](E_\alpha[(y_{n+1}^r)^2] - E_\alpha[y_{n+1}^r]^2) - \sum_{\alpha=1}^m E_j[x_{n+1}^\alpha] E_\alpha[y_{n+1}^r]\,.$$

This can be expressed alternatively in the form

$$(\mathbf{C}_Z^j)_{rr} - a_{rj} = \sum_{\alpha=1}^m \sum_{\beta=1}^m t_{r\alpha} (\mathbf{C}_X^j)_{\alpha\beta} t_{r\beta} + \sum_{\alpha=1}^m b_{\alpha j} (\mathbf{V}_Y^\alpha)_{rr} - \sum_{\alpha=1}^m b_{\alpha j} t_{r\alpha}\,.$$

Note that the negative terms on both sides are equal, since $\sum_{\alpha=1}^{m} b_{\alpha j} t_{r\alpha} = (\mathbf{TB})_{rj} = a_{rj}$. Then, from (i) and (ii), we obtain $\mathbf{C}_Z^j = \mathbf{TC}_X^j\mathbf{T}^{\mathrm{T}} + \sum_{\alpha=1}^{N_i} b_{\alpha j}\mathbf{V}_Y^{\alpha}$ as we wanted to show. The rest of the result is immediate by the definition of $\mathbf{V}_Z^j$. □

As a result of the discussion above, the global process is a g.w.p. $\mathbf{X}_0, \mathbf{X}_1, \ldots$ that can be considered as the composition of k iterations of the fast process followed by one iteration of the slow process. Moreover, its p.g.f. is the function $\mathbf{G}_k : \mathbb{R}^N \to \mathbb{R}^N$

$$\mathbf{G}_k(\mathbf{s}) = (G_k^{11}(\mathbf{s}), \ldots, G_k^{1N_1}(\mathbf{s}), G_k^{21}(\mathbf{s}), \ldots, G_k^{2N_2}(\mathbf{s}), \ldots, G_k^{q1}(\mathbf{s}), \ldots, G_k^{qN_q}(\mathbf{s}))\,,$$

given by

$$\mathbf{G}_k(\mathbf{s}) := \mathbf{G}_F \circ \mathbf{G}_F \circ \overset{(k)}{\cdots} \circ \mathbf{G}_F \circ \mathbf{G}_S(\mathbf{s}) = \mathbf{G}_{F,k} \circ \mathbf{G}_S(\mathbf{s})\,, \tag{4.3}$$

where

$$\mathbf{G}_{F,k} := \mathbf{G}_F \circ \mathbf{G}_F \circ \overset{(k)}{\cdots} \circ \mathbf{G}_F\,.$$

From (4.2) we have

$$G_{F,k}^{ij}(\mathbf{s}) = \sum_{\alpha} pr(\mathbf{w}_{t+k} = \alpha \,|\, \mathbf{w}_t = \mathbf{e}^{ij})\mathbf{s}^{\alpha} = \sum_{h=1}^{N_i} (\mathbf{P}_i^k)_{hj} s_{ih}\,,$$

$i = 1, \ldots, q; j = 1, \ldots, N_i$. So the expression of the p.g.f. of the global system in terms of the characteristics of the slow and fast processes is

$$G_k^{ij}(\mathbf{s}) = \sum_{\alpha} p_k^{ij}(\alpha)\mathbf{s}^{\alpha} = \sum_{\alpha}\sum_{h=1}^{N_i} p_S^{ih}(\alpha)(\mathbf{P}_i^k)_{hj}\mathbf{s}^{\alpha}\,, \quad i = 1, \ldots, q\,, \quad j = 1, \ldots, N_i\,.$$

Therefore, the transition probabilities corresponding to the original system are given by

$$p_k^{ij}(\alpha) = pr(\mathbf{X}_{n+1} = \alpha \,|\, \mathbf{X}_n = \mathbf{e}^{ij}) = \sum_{h=1}^{N_i} p_S^{ih}(\alpha)(\mathbf{P}_i^k)_{hj}\,.$$

Using Lemma 1, we have that the matrix of expected values for the original system is $\mathbf{MP}^k$.

5. Approximate Reduction of the Global Model

5.1. *The auxiliary model*

From the previous section we have that the fast process corresponds to a positive Markov chain defined, for each group i, by a primitive column stochastic matrix $\mathbf{P}_i$. Therefore, the fast process in each group i has a stationary probability distribution given by the positive right eigenvector $\mathbf{v}_i = (v_i^1, \ldots, v_i^{N_i})^{\mathrm{T}}$ of matrix $\mathbf{P}_i$ associated to eigenvalue 1 and normalized so that the sum of its components is one, i.e.

$$\mathbf{P}_i\mathbf{v}_i = \mathbf{v}_i\,, \quad \mathbf{1}_i^{\mathrm{T}}\mathbf{v}_i = 1\,,$$

where $\mathbf{1}_i = (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^{N_i \times 1}$. Note that, since each $\mathbf{P}_i$ is stochastic $\mathbf{1}_i^{\mathrm{T}}\mathbf{P}_i = \mathbf{1}_i^{\mathrm{T}}$.

Vector $\mathbf{v}_i$ may be interpreted in terms of the behavior of the fast process in group i. Consider the hypothetical situation in which the system was governed by the fast process exclusively, and assume that Δ_n is long enough with respect to the projection interval corresponding to the fast process for this to reach its equilibrium conditions during Δ_n. In the sequel we will refer to this situation of equilibrium for the fast process as "equilibrium fast process". Then, for a population at time n consisting of one individual in state (i, l), we would have that at the end of Δ_n, the probability of the individual being in state (i, j) would be v_i^j (which, moreover, is independent of l).

In this way, for each $i = 1, \ldots, q$, the matrix that characterizes the equilibrium probability distribution for the fast process in group i is

$$\bar{\mathbf{P}}_i = \lim_{k\to\infty} \mathbf{P}_i^k = \mathbf{v}_i \mathbf{1}_i^{\mathrm{T}} > 0\,,$$

and, therefore, the equilibrium conditions for the fast dynamics in the whole population are characterized by matrix $\bar{\mathbf{P}} = \mathrm{diag}(\bar{\mathbf{P}}_1, \bar{\mathbf{P}}_2, \ldots, \bar{\mathbf{P}}_q)$.

Let $\mathbf{G}_{\bar{F}}$ be the p.g.f. for the equilibrium fast process, i.e.

$$\mathbf{G}_{\bar{F}} := \lim_{k\to\infty} \mathbf{G}_{F,k}\,. \tag{5.4}$$

Then, its (ij)th component is

$$\begin{aligned} G_{\bar{F}}^{ij}(\mathbf{s}) &= \lim_{k\to\infty} G_{F,k}^{ij}(\mathbf{s}) = \sum_{h=1}^{N_i} \lim_{k\to\infty} (\mathbf{P}_i^k)_{hj} s_{ih} \\ &= \sum_{h=1}^{N_i} v_i^h s_{i\alpha}\,, \quad i = 1, \ldots, q\,; \quad j = 1, \ldots, N_i\,. \end{aligned} \tag{5.5}$$

Note that $G_{\bar{F}}^{ij}$ is independent of j, so that for each $i = 1, \ldots, q$, we can define $G_{\bar{F}}^{i} := G_{\bar{F}}^{ij} = \lim_{k\to\infty} G_{F,k}^{ij}$ for all $j = 1, \ldots, N_i$ and then

$$\mathbf{G}_{\bar{F}}(\mathbf{s}) = (G_{\bar{F}}^1(\mathbf{s}), \overset{(N_1)}{\ldots}, G_{\bar{F}}^1(\mathbf{s}), G_{\bar{F}}^2(\mathbf{s}), \overset{(N_2)}{\ldots}, G_{\bar{F}}^2(\mathbf{s}), \ldots, G_{\bar{F}}^q(\mathbf{s}), \overset{(N_q)}{\ldots}, G_{\bar{F}}^q(\mathbf{s}))\,, \tag{5.6}$$

Moreover, we define matrices

$$\mathbf{V} := \mathrm{diag}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_q)\,, \quad \mathbf{U} := \mathrm{diag}(\mathbf{1}_1^{\mathrm{T}}, \mathbf{1}_2^{\mathrm{T}}, \ldots, \mathbf{1}_q^{\mathrm{T}})\,.$$

Some of the properties of these matrices are gathered in the following lemma, whose proof is straightforward:

Lemma 2. *Matrices* $\mathbf{P}, \bar{\mathbf{P}}, \mathbf{V}$ *and* $\mathbf{U}$ *verify*:

(a) $\mathbf{P}\bar{\mathbf{P}} = \bar{\mathbf{P}}\mathbf{P} = \bar{\mathbf{P}}$,
(b) $\mathbf{P}\mathbf{V} = \mathbf{V}$,
(c) $\mathbf{U}\bar{\mathbf{P}} = \mathbf{U}$, $\mathbf{U}\mathbf{V} = \mathbf{I}_q$, $\bar{\mathbf{P}} = \mathbf{V}\mathbf{U}$.

We define the auxiliary process as the g.w.p. corresponding to the original system when the fast process has reached equilibrium, i.e. the system $\mathbf{X}'_0, \mathbf{X}'_1, \mathbf{X}'_2, \dots$ with $\mathbf{X}'_0 = \mathbf{X}_0$ whose p.g.f. is

$$\mathbf{G}'(\mathbf{s}) := \lim_{k\to\infty} \mathbf{G}_k(\mathbf{s}) = \lim_{k\to\infty} \mathbf{G}_{F,k} \circ \mathbf{G}_S(\mathbf{s}) = \mathbf{G}_{\bar{F}} \circ \mathbf{G}_S(\mathbf{s}) . \tag{5.7}$$

Taking into account (5.6) we have that the (ij)th component of $\mathbf{G}'$ is independent of j and then

$$\mathbf{G}'(\mathbf{s}) = (G'^1(\mathbf{s}), \overset{(N_1)}{\dots}, G'^1(\mathbf{s}), G'^2(\mathbf{s}), \overset{(N_2)}{\dots}, G'^2(\mathbf{s}), \dots, G'^q(\mathbf{s}), \overset{(N_q)}{\dots}, G'^q(\mathbf{s})) ,$$

where

$$G'^i(\mathbf{s}) = \sum_{h=1}^{N_i} v_i^h G_S^{ih}(\mathbf{s}) = \sum_\alpha \sum_{h=1}^{N_i} p_S^{ih}(\alpha) v_i^h \mathbf{s}^\alpha , \quad i = 1, \dots, q . \tag{5.8}$$

Note that for an initial population consisting in one individual, the dynamics of the auxiliary system will depend on the group but not the subgroup to which this individual belongs. The transition probabilities for the auxiliary system are

$$p'^i(\alpha) := pr(\mathbf{X}'_{n+1} = \alpha \,|\, \mathbf{X}'_n = \mathbf{e}^{ij}) = \sum_{h=1}^{N_i} p_S^{ih}(\alpha) v_i^h ,$$
$$i = 1, \dots, q , \quad j = 1, \dots, N_i , \tag{5.9}$$

not depending on the subgroup of the parent.

Using Lemma 1(b) we have that the matrix of expected values for the auxiliary system is $\mathbf{M}\bar{\mathbf{P}}$. Besides, a straightforward calculation from Lemma 1(c) shows that the covariance matrices of offspring production for the auxiliary system are given by

$$\mathbf{V}'^j = \sum_{l=1}^{N_j} v_j^l \mathbf{V}_S^{jl} + \mathbf{W}^j , \quad j = 1, \dots, q , \tag{5.10}$$

where $\mathbf{W}^j \in \mathbb{R}^{N\times N}$ is a symmetric matrix with q^2 blocks $\mathbf{W}^j_{rs} \in \mathbb{R}^{N_r \times N_s}$ given by

$$\mathbf{W}^j_{rs} = \mathbf{M}_{rj}(\mathrm{diag}(\mathbf{v}_j) - \mathbf{v}_j \mathbf{v}_j^{\mathrm{T}})\mathbf{M}_{sj}^{\mathrm{T}} . \tag{5.11}$$

5.2. *Aggregated system*

We define the macrovariables as the sum of the variables corresponding to each of the groups in the auxiliary system

$$y_n^i = x_n'^{i1} + x_n'^{i2} + \cdots + x_n'^{iN_i} , \quad i = 1, \dots, q ,$$

and therefore the vector of macrovariables for the population is

$$\mathbf{Y}_n = (y_n^1, y_n^2, \dots, y_n^q)^{\mathrm{T}} \in \mathbb{R}^{q\times 1}$$

so we can write $\mathbf{Y}_n = \mathbf{U}\mathbf{X}'_n$.

For each $i = 1, \ldots, q$, let $\mathbf{e}^i$ be the ith canonical vector of $\mathbb{R}^q$ and let the p.g.f. of the aggregated system be $\bar{\mathbf{G}} : \mathbb{R}^q \to \mathbb{R}^q$. Then $\bar{\mathbf{G}}(\mathbf{s}) = (\bar{G}^1(\mathbf{s}), \bar{G}^2(\mathbf{s}), \ldots, \bar{G}^q(\mathbf{s}))$ where

$$\bar{G}^i(\mathbf{s}) = \sum_{\beta} pr(\mathbf{Y}_{n+1} = \beta \,|\, \mathbf{Y}_n = \mathbf{e}^i) s_1^{\beta_1} \cdots s_q^{\beta_q}\,, \quad i = 1, \ldots, q\,,$$

being the summation extended to all $\beta \in \mathbb{Z}_+^q$. Note that due to the hypothesis on the slow dynamics, $pr(\mathbf{Y}_{n+1} = \beta \,|\, \mathbf{Y}_n = \mathbf{e}^i)$ will be null but for a finite number of vectors β.

The following result allows one to obtain the p.g.f. of the aggregated system in terms of the characteristics of the fast and the slow processes:

Proposition 1. *The p.g.f. of the aggregated system verifies*

$$\bar{G}^i(t_1, t_2, \ldots, t_q) = G'^i(t_1, \overset{(N_1)}{\ldots}, t_1, t_2, \overset{(N_2)}{\ldots}, t_2, t_q, \overset{(N_q)}{\ldots}, t_q)\,,$$

or, in abbreviated notation,

$$\bar{G}^i(\mathbf{t}) = G'^i(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_q) = \sum_{h=1}^{N_i} v_i^h G_S^{ih}(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_q)\,, \quad i = 1, \ldots, q\,, \tag{5.12}$$

where $\mathbf{t} = (t_1, t_2, \ldots, t_q)$ *and* $\mathbf{t}_i = (t_i, \overset{(N_i)}{\ldots}, t_i)$.

Proof. For each $i = 1, \ldots, q$ and $r = 1, \ldots, N_i$ we know that G'^{ir} does not depend on r. Then we have, by definition, $G'^i(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_q) = \sum_\alpha pr(\mathbf{X}'_{n+1} = \alpha \,|\, \mathbf{X}'_n = \mathbf{e}^{ir}) t_1^{\alpha_{11}+\cdots+\alpha_{1N_1}} \ldots t_q^{\alpha_{q1}+\cdots+\alpha_{qN_q}}$ where $pr(\mathbf{X}'_{n+1} = \alpha \,|\, \mathbf{X}'_n = \mathbf{e}^{ir})$ is independent of r $(*)$. For $i = 1, \ldots, q$ let $\alpha_i = \alpha_{i1} + \alpha_{i2} + \cdots + \alpha_{iN_i}$. Now, we sum over the α_{is} in two stages, first keeping α_i constant and then summing over all the possible values of α_i. In this way, we have

$$\begin{aligned} &G'^i(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_q) \\ &\quad = \sum_{\beta_1,\ldots,\beta_q} \sum_{\alpha_1=\beta_1} \cdots \sum_{\alpha_q=\beta_q} pr(\mathbf{X}'_{n+1} = \alpha \,|\, \mathbf{X}'_n = \mathbf{e}^{ir}) t_1^{\beta_1} \cdots t_q^{\beta_q}\,. \end{aligned} \tag{5.13}$$

In order to shorten the notation, let us denote $\xi^i(\beta) = \sum_{\alpha_1=\beta_1} \cdots \sum_{\alpha_q=\beta_q} pr(\mathbf{X}'_{n+1} = \alpha \,|\, \mathbf{X}'_n = \mathbf{e}^{ir})$ and $\rho^i(\beta) = pr(\mathbf{Y}_{n+1} = \beta \,|\, \mathbf{Y}_n = \mathbf{e}^i)$. All we need to prove now is that $\xi^i(\beta) = \rho^i(\beta)$ for each β. In that case, we would have from (5.13) $G'^i(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_q) = \sum_{\beta_1,\ldots,\beta_q} pr(\mathbf{Y}_{n+1} = \beta \,|\, \mathbf{Y}_n = \mathbf{e}^i) t_1^{\beta_1} \cdots t_q^{\beta_q} = \bar{G}^i(t_1, t_2, \ldots, t_q)$, as we want to show. Now, by the definition of the macrovariables,

$$\begin{aligned} \rho^i(\beta) &= pr\left(\sum_{l=1}^{N_j} x'^{jl}_{n+1} = \beta_j; j = 1, \ldots, q \,|\, \mathbf{Y}_n = \mathbf{e}^i\right) \\ &= pr\left(\sum_{l=1}^{N_j} x'^{jl}_{n+1} = \beta_j; j = 1, \ldots, q \,|\, \bigcup_{r=1}^{N_i} (\mathbf{X}'_n = \mathbf{e}^{ir})\right) \end{aligned}$$

$$= \left(\sum_{r=1}^{N_i} pr(\mathbf{X}'_n = \mathbf{e}^{ir}) \right)^{-1} \sum_{r=1}^{N_i} pr \left(\sum_{l=1}^{N_j} x'^{jl}_{n+1} = \beta_j; j = 1, \ldots, q \,|\, \mathbf{X}'_n = \mathbf{e}^{ir} \right)$$

$$\times\, pr(\mathbf{X}'_n = \mathbf{e}^{ir}) \,,$$

where in the second equality we have expressed the event $(\mathbf{Y}_n = \mathbf{e}^i)$ as union of the disjoint events $(\mathbf{X}'_n = \mathbf{e}^{ir}); r = 1, \ldots, N_i$. Now we write $(\sum_{l=1}^{N_j} x'^{jl}_{n+1} = \beta_j; j = 1, \ldots, q) = \bigcup_{\alpha_{11}+\cdots+\alpha_{1N_1}=\beta_1} \cdots \bigcup_{\alpha_{q1}+\cdots+\alpha_{qN_q}=\beta_q} (\mathbf{X}'_{n+1} = \alpha)$, where the union is over all the possible vectors $\alpha \in \mathbb{Z}_+^N$ such that $\alpha_{i1}+\cdots+\alpha_{iN_i} = \beta_i$ for all $i = 1, \ldots, q$. Then we have $\rho^i(\beta) = (\sum_{r=1}^{N_i} pr(\mathbf{X}'_n = \mathbf{e}^{ir}))^{-1} \sum_{r=1}^{N_i} \sum_{\alpha_1=\beta_1} \cdots \sum_{\alpha_q=\beta_q} pr(\mathbf{X}'_{n+1} = \alpha \,|\, \mathbf{X}'_n = \mathbf{e}^{ir}) pr(\mathbf{X}'_n = \mathbf{e}^{ir}) = (\sum_{r=1}^{N_i} pr(\mathbf{X}'_n = \mathbf{e}^{ir}))^{-1} \sum_{\alpha_1=\beta_1} \cdots \sum_{\alpha_q=\beta_q} pr(\mathbf{X}'_{n+1} = \alpha \,|\, \mathbf{X}'_n = \mathbf{e}^{ir}) \sum_{r=1}^{N_i} pr(\mathbf{X}'_n = \mathbf{e}^{ir}) = \xi^i(\beta)$ as we wanted to show, having used $(*)$ in the second equality. □

Using the previous result we can obtain the matrices of first and second order moments for offspring production in the aggregated system.

Proposition 2. (a) *The matrix* $\bar{\mathbf{M}} \in \mathbb{R}^{q\times q}$ *of expected values for the aggregated system is*

$$\bar{\mathbf{M}} = \mathbf{UMV} \in \mathbb{R}^{q\times q} \,. \tag{5.14}$$

(b) *Moreover, the covariance matrices* $\bar{\mathbf{V}}^j \in \mathbb{R}^{q\times q}$ *for offspring production in the aggregated system* (*where* $(\bar{\mathbf{V}}^j)_{rl} = \mathrm{Cov}(y^r_{n+1}, y^l_{n+1} \,|\, \mathbf{Y}_n = \mathbf{e}^j)$ *for each* $r, l = 1, \ldots, q$) *are given by*

$$\bar{\mathbf{V}}^j = \mathbf{U}\mathbf{V}'^j\mathbf{U}^{\mathrm{T}} = \sum_{l=1}^{N_i} v^l_j \mathbf{U}\mathbf{V}^{jl}_S \mathbf{U}^{\mathrm{Ts}} + \mathbf{H}^j \,, \quad j = 1, \ldots, q$$

where the block (r, s) *of* $\mathbf{H}^j$ *is* $\mathbf{H}^j_{rs} = \mathbf{1}_r^{\mathrm{T}}\mathbf{M}_{rj}(\mathrm{diag}(\mathbf{v}_j) - \mathbf{v}_j\mathbf{v}_j^{\mathrm{T}})\mathbf{M}^{\mathrm{T}}_{sj}\mathbf{1}_s$.

Proof. The results follow by differentiation in (5.12). (a) The first derivatives of the p.g.f. associated to the aggregated system are

$$\frac{\partial \bar{G}^j}{\partial t_i}(t_1, \ldots, t_q) = \frac{\partial G'^j(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_q)}{\partial t_i} = \sum_{l=1}^{N_j} v^l_j \frac{\partial G^{jl}_S}{\partial t_i}(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_q)$$

$$= \sum_{l=1}^{N_j} v^l_j \sum_{r=1}^{N_i} \frac{\partial G^{jl}_S}{\partial s_{ir}}(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_q) \,, \quad j = 1, \ldots, q \,.$$

Let $\bar{\mathbf{M}} = [\bar{m}_{ij}]$. Then

$$\bar{m}_{ij} = E_j[y^i_{n+1}] = \frac{\partial \bar{G}^j}{\partial t_i}(\mathbf{1}) = \sum_{l=1}^{N_j} v^l_j \sum_{r=1}^{N_i} \frac{\partial G^{jl}_S}{\partial s_{ir}}(\mathbf{1}) = \sum_{l=1}^{N_j} v^l_j \sum_{r=1}^{N_i} M^{rl}_{ij} = \mathbf{1}_i^{\mathrm{T}}\mathbf{M}_{ij}\mathbf{v}_j \,,$$

as we wanted to show. (b) The second derivatives of $\bar{G}^j$ are $\frac{\partial^2 \bar{G}^j}{\partial t_h \partial t_i}(t_1, \ldots, t_q) = \sum_{r=1}^{N_i} \sum_{s=1}^{N_h} \frac{\partial^2 G'^j(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_q)}{\partial s_{hs} \partial s_{ir}}$; $i, j, h = 1, \ldots, q$. Now, we distinguish two cases: (i) If $i \neq h$ we have

$$E_j(y_{n+1}^i y_{n+1}^h) = \sum_{r=1}^{N_i} \sum_{s=1}^{N_h} E_{jl}[x_{n+1}'^{ir} x_{n+1}'^{hs}] \; (*)$$

for any $l = 1, \ldots, N_j$. (ii) If $i = h$, then

$$\begin{aligned} E_j[(y_{n+1}^i)^2] - E_j[y_{n+1}^i] &= \sum_{r=1}^{N_i} \sum_{s=1, s \neq r}^{N_i} E_{jl}[x_{n+1}'^{ir} x_{n+1}'^{is}] \\ &\quad + \sum_{r=1}^{N_i} (E_{jl}[(x_{n+1}'^{ir})^2] - E_{jl}[x_{n+1}'^{ir}]) \\ &= \sum_{r=1}^{N_i} \sum_{s=1}^{N_i} E_{jl}[x_{n+1}'^{ir} x_{n+1}'^{is}] - \sum_{r=1}^{N_i} E_{jl}[x_{n+1}'^{ir}] \,. \end{aligned}$$

Note that the negative terms on both sides are equal, since

$$\sum_{r=1}^{N_i} E_{jl}[x_{n+1}'^{ir}] = E_{jl}\left[\sum_{r=1}^{N_i} x_{n+1}'^{ir}\right] = E_j[y_{n+1}^i] \; (**)$$

and so $(*)$ holds everywhere. On the other hand, from $(**)$ we have

$$\begin{aligned} E_j[y_{n+1}^i] E_j[y_{n+1}^h] &= \left(\sum_{r=1}^{N_i} E_{jl}[x_{n+1}'^{ir}]\right) \left(\sum_{s=1}^{N_h} E_{jl}[x_{n+1}'^{hs}]\right) \\ &= \sum_{r=1}^{N_i} \sum_{s=1}^{N_h} E_{jl}[x_{n+1}'^{ir}] E_{jl}[x_{n+1}'^{hs}] \,. \end{aligned}$$

From this expression and $(*)$ it follows that

$$\mathrm{Cov}(y_{n+1}^i y_{n+1}^h \,|\, \mathbf{Y}_n = \mathbf{e}^j) = \sum_{r=1}^{N_i} \sum_{s=1}^{N_h} \mathrm{Cov}(x_{n+1}'^{ir} x_{n+1}'^{hs} \,|\, \mathbf{X}_n' = \mathbf{e}^{jl}) \,,$$

for every $i, h = 1, \ldots, q$, (where the value of l is irrelevant) which can be expressed in matrix form as $\bar{\mathbf{V}}^j = \mathbf{U}\mathbf{V}'^j\mathbf{U}^{\mathrm{T}}$, being $\bar{\mathbf{V}}^j$ and $\mathbf{V}'^j$, respectively, the covariance matrices for offspring production of the aggregated and the auxiliary systems given a parent in state j. Now, the result is a direct consequence of (5.10) and (5.11). $\square$

5.3. *Aggregation of a multiregional model with fast migration*

Next we will build an aggregated system for the multiregional model described in Sec. 3, according to the procedure proposed in Sec. 5. First we must remark that, in this example, the migration process satisfies hypotheses (HA) and (HB) established in Sec. 4 for the fast process. Furthermore, the stationary probability distribution of the migration process for each group i is given by v_i, the normalized right eigenvector associated to eigenvalue 1 in matrix P_i (note that these matrices were assumed to be column stochastic and primitive). This vector can be obtained from the transition probabilities of the process as

$$\mathbf{v}_i = (v_i^1, v_i^2)^{\mathrm{T}} = \left(\frac{p_F^{i2}(\mathbf{e}^{i1})}{p_F^{i1}(\mathbf{e}^{i2}) + p_F^{i2}(\mathbf{e}^{i1})}, \frac{p_F^{i1}(\mathbf{e}^{i2})}{p_F^{i1}(\mathbf{e}^{i2}) + p_F^{i2}(\mathbf{e}^{i1})}\right)^{\mathrm{T}}.$$

Now, from Eq. (5.8), the auxiliary process $\mathbf{X}_n' = (x_n'^{11}, x_n'^{12}, x_n'^{21}, x_n'^{22})^{\mathrm{T}}$ has p.g.f.

$$\mathbf{G}'(\mathbf{s}) = (G'^1(\mathbf{s}), G'^1(\mathbf{s}), G'^2(\mathbf{s}), G'^2(\mathbf{s}))$$

where $G'^1(\mathbf{s}) = v_1^1 G_S^{11}(\mathbf{s}) + v_1^2 G_S^{12}(\mathbf{s})$, $G'^2(\mathbf{s}) = v_2^1 G_S^{21}(\mathbf{s}) + v_2^2 G_S^{22}(\mathbf{s})$, and its matrix of expected values is $\mathbf{M}\bar{\mathbf{P}}$.

The two variables corresponding to the reduced system are the total population in each class for the auxiliary system, i.e.

$$y_n^i = x_n'^{i1} + x_n'^{i2}, \quad i = 1, 2$$

and the p.g.f. of the aggregated system is obtained by applying Proposition 1

$$\begin{aligned}\bar{G}^1(s_1, s_2) &= G'^1(s_1, s_1, s_2, s_2) = \bar{p}^1(0,0) + \bar{p}^1(1,0)s_1 + \bar{p}^1(2,0)s_1^2 \\ &\quad + \bar{p}^1(0,1)s_2 + \bar{p}^1(1,1)s_1 s_2 + \bar{p}^1(2,1)s_1^2 s_2, \\ \bar{G}^2(s_1, s_2) &= G'^2(s_1, s_1, s_2, s_2) = \bar{p}^2(0,0) + \bar{p}^2(1,0)s_1 + \bar{p}^2(2,0)s_1^2,\end{aligned}$$

where

$$\begin{aligned}&\bar{p}^1(0,0) = v_1^1 p_{0000}^{11} + v_1^2 p_{0000}^{12}, \quad \bar{p}^1(1,0) = v_1^1 p_{1000}^{11} + v_1^2 p_{0100}^{12}, \\ &\bar{p}^1(2,0) = v_1^1 p_{2000}^{11} + v_1^2 p_{0200}^{12}, \quad \bar{p}^1(0,1) = v_1^1 p_{0010}^{11} + v_1^2 p_{0001}^{12}, \\ &\bar{p}^1(1,1) = v_1^1 p_{1010}^{11} + v_1^2 p_{0101}^{12}, \quad \bar{p}^1(2,1) = v_1^1 p_{2010}^{11} + v_1^2 p_{0201}^{12}, \\ &\bar{p}^2(0,0) = v_2^1 p_{0000}^{21} + v_2^2 p_{0000}^{22}, \bar{p}^2(1,0) = v_2^1 p_{1000}^{21} + v_2^2 p_{0100}^{22}, \\ &\bar{p}^2(2,0) = v_2^1 p_{2000}^{21} + v_2^2 p_{0200}^{22}.\end{aligned}$$

Finally, the matrix of expectations for the aggregated system is

$$\bar{\mathbf{M}} = \mathbf{UMV} = \begin{pmatrix} m_{11}v_{11} + m_{22}v_{12} & m_{13}v_{21} + m_{24}v_{22} \\ m_{31}v_{11} + m_{42}v_{12} & 0 \end{pmatrix}.$$

In summary, our technique has taken advantage of the presence of time scales to reduce a g.w.p. with four types to an aggregated system with only two types. The reduced system is constructed in terms of the probabilities defining demography and the stationary equilibrium distribution of migration. Note that if we consider a multiregional model like the one above but with q age classes and t spatial patches, we would have to deal with a g.w.p. with qt types. Our aggregation procedure renders a reduced model with only q types.

6. Relationships Between the Original and the Aggregated System

In order to relate the spectral properties of the matrices $\mathbf{M}\bar{\mathbf{P}}$ and $\bar{\mathbf{M}}$ of expected values for the auxiliary and aggregated systems respectively, we will make use of the following lemma.

Lemma 3. *Matrices* $\mathbf{M}\bar{\mathbf{P}}$ *and* $\bar{\mathbf{M}}$ *verify*:

(a) *For* $n \geq 2$,

$$(\mathbf{M}\bar{\mathbf{P}})^n = \mathbf{M}\mathbf{V}\bar{\mathbf{M}}^{n-1}\mathbf{U}\,, \quad \bar{\mathbf{M}}^n = \mathbf{U}(\mathbf{M}\bar{\mathbf{P}})^{n-1}\mathbf{M}\mathbf{V}\,.$$

(b) $\det(\lambda \mathbf{I}_N - \mathbf{M}\bar{\mathbf{P}}) = \lambda^{N-q}\det(\lambda \mathbf{I}_q - \bar{\mathbf{M}})$; *in particular*, *the dominant eigenvalues of matrices* $\mathbf{M}\bar{\mathbf{P}}$ *and* $\bar{\mathbf{M}}$, *together with their respective multiplicities*, *coincide.*

(c) *If* $\mathbf{r}$ *and* $\mathbf{l}$ *are respectively right and left eigenvectors of* $\bar{\mathbf{M}}$ *associated to* $\lambda \neq 0$ *then* $\mathbf{M}\mathbf{V}\mathbf{r}$ *and* $\mathbf{U}^{\mathrm{T}}\mathbf{l}$ *are respectively right and left eigenvectors of* $\mathbf{M}\bar{\mathbf{P}}$ *associated to* λ.

Proof. (a) Straightforward from (5.14) and Lemma 2. (b) Direct consequence of the fact that the nonzero eigenvalues, including multiplicities, of $\mathbf{AB}$ coincide with those of $\mathbf{BA}$.[11] (c) We know $\bar{\mathbf{M}}\mathbf{r} = \lambda\mathbf{r} \neq \mathbf{0}$, i.e. $\mathbf{UMVr} = \lambda\mathbf{r} \neq \mathbf{0}$ $(*)$ so it must be $\mathbf{MVr} \neq \mathbf{0}$. Multiplying on the left by $\mathbf{MV}$ and using Lemma 2 we have $\mathbf{M}\bar{\mathbf{P}}\mathbf{MVr} = \lambda\mathbf{MVr}$. Analogously, we know that $\mathbf{l}^{\mathrm{T}}\bar{\mathbf{M}} = \lambda\mathbf{l}^{\mathrm{T}} \neq \mathbf{0}$, i.e. $\mathbf{l}^{\mathrm{T}}\mathbf{UMV} = \lambda\mathbf{l}^{\mathrm{T}} \neq \mathbf{0}$, so $\mathbf{l}^{\mathrm{T}}\mathbf{U} \neq \mathbf{0}$. Multiplying on the right by $\mathbf{U}$ and using Lemma 2 we have $\mathbf{l}^{\mathrm{T}}\mathbf{U}\mathbf{M}\bar{\mathbf{P}} = \lambda\mathbf{l}^{\mathrm{T}}\mathbf{U}$, as we wanted to show. □

From hypothesis (HB) on the fast process, the blocks $\mathbf{P}_i$ are column-stochastic primitive matrices, so they have 1 as their dominant eigenvalue which, moreover, is simple and the subdominant eigenvalues have modulus strictly lower than 1. Now let us consider the eigenvalues of $\mathbf{P}$ ordered by decreasing modulus (note that the set of eigenvalues of $\mathbf{P}$ is the union of those corresponding to different $\mathbf{P}_i$)

$$1 = \gamma_1 = \gamma_2 = \cdots = \gamma_q > |\gamma_{q+1}| \geq \cdots \geq |\gamma_N|\,,$$

and let

$$\gamma > |\gamma_{q+1}|\,, \tag{6.15}$$

i.e. γ is any real number greater than the modulus of the greater "subdominant eigenvalue" of $\mathbf{P}$. Note that γ can always be taken smaller than 1.

Let us introduce some concepts and notation which will be useful in the subsequent developments. A non-negative matrix $\mathbf{A}$ is said to be column allowable (row allowable) if it has at least a nonzero element in each of its columns (rows). $\mathbf{A}$ is said to be allowable if it is both column and row allowable. The product of row (column) allowable matrices is row (column) allowable. It is easy to check that if $\mathbf{A}$ is row (column) allowable and $\mathbf{B}$ is a positive matrix, then $\mathbf{AB} > \mathbf{0}$ ($\mathbf{BA} > \mathbf{0}$) as long as the product is defined. If $\mathbf{A} = [a_{ij}]$ is any real matrix, we will denote by $|\mathbf{A}|$ the matrix $[|a_{ij}|]$ where $|*|$ denotes the absolute value.

In the following we will say that a property holds "for large enough k" when there exists an integer k_0 such that the property holds for $k \geq k_0$.

The following proposition, which in particular relates some of the spectral properties of matrices $\mathbf{MP}^k$ and $\bar{\mathbf{M}}$, hinges on results obtained by the authors in Ref. 17 to study aggregation techniques in a deterministic context:

Proposition 3. (a) *The matrices of expected values for the original system and the auxiliary model are related by* $\mathbf{MP}^k = \mathbf{M}\bar{\mathbf{P}} + \mathbf{o}(\gamma^k)$*;* $k \to \infty$. *In particular, for* k *large enough, both matrices have the same incidence matrix.*

(b) *Moreover, let* λ *be a simple and strictly dominant eigenvalue of* $\bar{\mathbf{M}}$ *associated to right and left eigenvectors* $\mathbf{r}$ *and* $\mathbf{l}$, *respectively. If* k *is large enough, matrix* $\mathbf{MP}^k$ *has a simple and strictly dominant eigenvalue* λ_k *that can be expressed in the form*

$$\lambda_k = \lambda + o(\gamma^k)\,,$$

and associated to λ_k *are right and left eigenvectors* $\mathbf{r}_k$ *and* $\mathbf{l}_k$ *that can be written in the form*

$$\mathbf{r}_k = \mathbf{MVr} + \mathbf{o}(\gamma^k)\,, \quad \mathbf{l}_k = \mathbf{U}^{\mathrm{T}}\mathbf{l} + \mathbf{o}(\gamma^k)\,.$$

Proof. Straightforward application of Propositions 4.4 and 4.5. in Ref. 17. Matrix $\bar{\mathbf{P}}$ here plays the role of matrix $\mathbf{A}$ in that reference. □

The properties of super/sub criticality, positive regularity and non-singularity[16] are related, for the original, auxiliary and aggregated systems, in the next result.

Proposition 4. *For large enough* k, *we have*:

(a) *If the original system is positively regular, then so is the aggregated system. Besides, if the aggregated system is positively regular and matrix* $\mathbf{M}$ *is row-allowable, then the original system is positively regular.*

(b) *The non-singularity of the original system is equivalent to the non-singularity of the auxiliary system and also equivalent to the non-singularity of the aggregated system.*

(c) *The original system is supercritical (subcritical) if and only if the aggregated system is supercritical (subcritical).*

Proof. (a) Using Proposition 3 we have that $\mathbf{MP}^k$ is primitive for large enough k if and only if $\mathbf{M}\bar{\mathbf{P}}$ is. Now assume $\mathbf{MP}^k$ (and consequently $\mathbf{M}\bar{\mathbf{P}}$) is primitive. This means that, for some n, $(\mathbf{M}\bar{\mathbf{P}})^n > 0$. Now, using Lemma 3(a) and the fact that matrices $\mathbf{V}$ and $\mathbf{U}$ are always allowable due to the positivity of vectors $\mathbf{v}_i$ and $\mathbf{1}_i$, we have $\bar{\mathbf{M}}^{n+1} = \mathbf{U}(\mathbf{M}\bar{\mathbf{P}})^n\mathbf{MV} > \mathbf{0}$ and so $\bar{\mathbf{M}}$ is primitive. Conversely, if $\bar{\mathbf{M}}$ is primitive, then $\bar{\mathbf{M}}^n > \mathbf{0}$ for some n. Now, if $\mathbf{M}$ is row-allowable, then so is $\mathbf{MV}$, and therefore, using Lemma 3, we have $(\mathbf{M}\bar{\mathbf{P}})^{n+1} = \mathbf{MV}\bar{\mathbf{M}}^n\mathbf{U} > \mathbf{0}$ which means that $\mathbf{M}\bar{P}$, and therefore $\mathbf{MP}^k$ for large enough k, is primitive.

(b) Recall that the singularity or non-singularity of a g.w.p. depends only on the signs (positive or zero) of the coefficients of its p.g.f. Using (5.4), (5.5) and the fact that $\mathbf{P}^k$ and $\bar{\mathbf{P}}$ have the same incidence matrix for large enough k, we have that for each value of α, the coefficient of $\mathbf{s}^\alpha$ in $\mathbf{G}_{\bar{F}}$ is nonzero if and only if the coefficient of $\mathbf{s}^\alpha$ in $\mathbf{G}_{F,k}$ is nonzero. Consequently, we have that the coefficient of $\mathbf{s}^\alpha$ in $\mathbf{G}'$ is nonzero if and only if the coefficient of $\mathbf{s}^\alpha$ in $\mathbf{G}_k$ is nonzero. Therefore, for large enough k the original system is non-singular if and only if the auxiliary system is. Now we will prove that the aggregated system is non-singular if and only if the auxiliary system is. Let $G'^i(\mathbf{s}) = \sum_\alpha p'^i(\alpha)s^\alpha$ for all $i = 1, \dots, q$ be the p.g.f. of the auxiliary system. Assume that the auxiliary system is singular, i.e. for all i, G'^i is a linear function. Then for all i and all $\alpha \in Z_+^N$ such that $p'^i(\alpha) > 0$ we have $\alpha_{11} + \cdots + \alpha_{1N_1} + \cdots + \alpha_{q1} + \cdots + \alpha_{qN_q} = 1$. Then $\bar{G}^i(\mathbf{t}) = \sum_\alpha p'^i(\alpha) t_1^{\alpha_{11}+\cdots+\alpha_{1N_1}} \cdots t_q^{\alpha_{q1}+\cdots+\alpha_{qN_q}}$ is a linear function for all $i = 1, \dots, q$, since one and only one of the α_{rs} is one and the rest is zero. Therefore the aggregated system is singular. Conversely, assume the auxiliary system is non-singular. Then there exists $i = 1, \dots, q$ and $\alpha \in Z_+^N$ such that $\alpha_{11} + \cdots + \alpha_{1N_1} + \cdots + \alpha_{q1} + \cdots + \alpha_{qN_q} \neq 1$ and $p'^i(\alpha) > 0$. Then $\bar{G}^i(\mathbf{t}) = \sum_\alpha p'^i(\alpha) t_1^{\alpha_{11}+\cdots+\alpha_{1N_1}} \cdots t_q^{\alpha_{q1}+\cdots+\alpha_{qN_q}}$ is not linear and so the aggregated system is non-singular.

(c) Direct consequence of Proposition 3(b). □

6.1. *Probability of extinction*

Let the probabilities of extinction by time n for the three systems be defined by

$$q_k^{ij}(n) = pr(\mathbf{X}_n = \mathbf{0} \,|\, \mathbf{X}_0 = \mathbf{e}^{ij})\,, \quad i = 1, \dots, q\,, \quad j = 1, \dots, N_i$$

$$q'^i(n) = q'^{ij}(n) = pr(\mathbf{X}'_n = \mathbf{0} \,|\, \mathbf{X}_0 = \mathbf{e}^{ij})\,, \quad i = 1, \dots, q\,, \quad j = 1, \dots, N_i$$

$$\bar{q}^i(n) = pr(\mathbf{Y}_n = \mathbf{0} \,|\, \mathbf{Y}_0 = \mathbf{e}^i)\,, \quad i = 1, \dots, q\,.$$

Note that, as a consequence of (5.9), $q'^{ij}(n)$ must be independent of j. Therefore, the vectors defining the probabilities of extinction by time n are given by

$$\mathbf{q}_k(n) = (q_k^{11}(n), \dots, q_k^{1N_1}(n), \dots, q_k^{q1}(n), \dots, q_k^{qN_q}(n)) \in \mathbb{R}^N\,,$$

$$\mathbf{q}'(n) = (q'^1(n), \overset{(N_1)}{\dots}, q'^1(n), \dots, q'^q(n), \overset{(N_q)}{\dots}, q'^q(n)) \in \mathbb{R}^N\,,$$

$$\bar{\mathbf{q}}(n) = (\bar{q}^1(n), \bar{q}^2(n), \dots, \bar{q}^q(n)) \in \mathbb{R}^q\,.$$

Since $\mathbf{Y}_n = \mathbf{U}\mathbf{X}'_n$ and $\mathbf{U}$ is an allowable matrix, it follows that for all $n = 1, 2, \dots,$

$$\mathbf{Y}_n = \mathbf{0} \text{ if and only if } \mathbf{X}'_n = \mathbf{0}\,, \tag{6.16}$$

i.e. the population goes extinct in the auxiliary system if and only if it goes extinct in the aggregated system.

The following result establishes, in a precise sense, a relationship between the original and the auxiliary system:

Lemma 4. *For all values of* $\alpha \in Z_+^N$:

(a) $p_k^{ij}(\alpha) = p'^i(\alpha) + o(\gamma^k)$,
(b) *For* n *fixed,* $pr(\mathbf{X}_n = \alpha \,|\, \mathbf{X}_0 = \mathbf{e}^{ij}) = pr(\mathbf{X}'_n = \alpha \,|\, \mathbf{X}_0 = \mathbf{e}^{ij}) + o(\gamma^k)$.
(c) $\mathbf{G}_k(\mathbf{s}) = \mathbf{G}'(\mathbf{s}) + \mathbf{o}(\gamma^k)$ *uniformly in* $0 \le \mathbf{s} \le 1$.

Proof. Absolutely analogous to that of Proposition 4 in Ref. 4. □

The probabilities of extinction by time n for the three systems are related by the following proposition.

Proposition 5. *For all* $n = 1, 2, \dots$

(a) $\mathbf{q}_k(n) = \mathbf{q}'(n) + \mathbf{o}(\gamma^k)$,
(b) $q'^i(n) = \bar{q}^i(n)\,, \quad i = 1, \dots, q$.
(c) $q_k^{ij}(n) = \bar{q}^i(n) + o(\gamma^k)\,, \quad i = 1, \dots, q\,;\ j = 1, \dots, N_i$.

Note that, from (c), the probabilities of extinction in finite time for the global process can be approximated by those corresponding to the aggregated process.

Proof. (a) Immediate consequence of Lemma 4 (b) for $\alpha = \mathbf{0}$. (b) Let $i = 1, \dots, q$ and $j = 1, \dots, N_i$ be fixed. Then

$$\bar{q}^i(n) = pr(\mathbf{Y}_n = \mathbf{0} \,|\, \mathbf{Y}_0 = \mathbf{e}^i) = pr\left(\mathbf{X}'_n = \mathbf{0} \,|\, \bigcup_{j=1}^{N_i} (\mathbf{X}'_0 = \mathbf{e}^{ij})\right)$$

$$= \left(\sum_{j=1}^{N_i} pr(\mathbf{X}'_0 = \mathbf{e}^{ij})\right)^{-1} \sum_{j=1}^{N_i} pr(\mathbf{X}'_n = \mathbf{0} \,|\, \mathbf{X}'_0 = \mathbf{e}^{ij}) pr(\mathbf{X}'_0 = \mathbf{e}^{ij})$$

$$= pr(\mathbf{X}'_n = \mathbf{0} \,|\, \mathbf{X}'_0 = \mathbf{e}^{ij}) = q'^i(n)\,,$$

where in the second equality we have used (6.16) and in the fourth the fact that $pr(\mathbf{X}'_n = \mathbf{0} \,|\, \mathbf{X}'_0 = \mathbf{e}^{ij})$ is independent of j. (c) Immediate from (a) and (b). □

The probabilities of ultimate extinction for the three systems are defined by

$$q_k^{ij} := pr\left(\lim_{n\to\infty} \mathbf{X}_n = 0 \,|\, \mathbf{X}_0 = \mathbf{e}^{ij}\right) = \lim_{n\to\infty} q_k^{ij}(n)\,, \quad i = 1, \dots, q\,; j = 1, \dots, N_i\,,$$

$$q'^i := pr\left(\lim_{n\to\infty} \mathbf{X}'_n = 0 \,|\, \mathbf{X}_0 = \mathbf{e}^{ij}\right) = \lim_{n\to\infty} q'^i(n)\,, \quad i = 1, \dots, q\,,$$

$$\bar{q}^i := pr\left(\lim_{n\to\infty} \mathbf{Y}_n = 0 \,|\, \mathbf{Y}_0 = \mathbf{e}^i\right) = \lim_{n\to\infty} \bar{q}^i(n)\,, \quad i = 1, \dots, q\,,$$

which can be gathered in the vectors $\mathbf{q}_k = (q_k^{11}, \dots, q_k^{1N_1}, \dots, q_k^{q1}, \dots, q_k^{qN_q})$, $\mathbf{q}' = (q'^1, \overset{(N_1)}{\dots}, q'^1, \dots, q'^q, \overset{(N_q)}{\dots}, q'^q)$ and $\bar{\mathbf{q}} = (\bar{q}^1, \bar{q}^2, \dots, \bar{q}^q)$. Using the fact that $q'^i(n) =$

$\bar{q}^i(n)$; $i = 1, \ldots, q$ and taking limits $n \to \infty$ we obtain

$$\bar{q}^i = q'^i\,, \quad i = 1, \ldots, q\,, \tag{6.17}$$

which relates the probabilities of ultimate extinction for the auxiliary process and the aggregated system. Regarding the relationship between the global process and the aggregated system we have the following proposition, which is a generalization of a similar result in Ref. 4:

Proposition 6. *Let the aggregated system be positively regular and non-singular and let* $\mathbf{M}$ *be row allowable* (*so the global system will also meet these requirements for large enough* k). *Then*:

(a) *If the aggregated system is subcritical, which implies* $\bar{\mathbf{q}} = \mathbf{1}$, *then the original system is also subcritical, and consequently* $\mathbf{q}_k = \mathbf{1}$.
(b) *If the aggregated system is supercritical, i.e.* $\bar{\mathbf{q}} < \mathbf{1}$, *then the original system is also supercritical, i.e.* $\mathbf{q}_k < \mathbf{1}$ *and besides*

$$\lim_{k\to\infty} q_k^{ij} = \bar{q}^i \quad \textit{for all } i = 1, \ldots, q\,, \quad j = 1, \ldots, N_i\,.$$

Proof. (a) Obvious from Proposition 4.
(b) Assume the aggregated process is supercritical. Then, Proposition 4 guarantees that the original system is also supercritical. Besides, using (6.17) we only need to show that $\lim_{k\to\infty} \mathbf{q}_k = \mathbf{q}'$. This has been done in Ref. 4 (Theorem 1), and all the reasonings carried out there, are valid in the present situation. □

6.2. *Moments of the population vector*

In order to study the statistical moments of the population vector of the global system in terms of those corresponding to the aggregated system, we introduce the following hypothesis:

H1. Matrix $\bar{\mathbf{M}}$ has a simple and strictly dominant eigenvalue λ associated to right and left eigenvectors $\mathbf{r}$ and $\mathbf{l}$ respectively, for which we will assume the following normalization conditions

$$\|\mathbf{r}\|_1 = 1\,, \quad \mathbf{l}^{\mathrm{T}}\mathbf{r} = 1\,. \tag{6.18}$$

Observe that hypothesis H1 is weaker than the condition that the aggregated system be positively regular.

6.2.1. *Expected values*

Note that if a certain square matrix $\mathbf{A}$ has a strictly dominant eigenvalue μ which is simple and is associated to right and left eigenvectors $\mathbf{v}$ and $\mathbf{u}$ respectively, we have

$$\lim_{n\to\infty} \frac{\mathbf{A}^n}{\mu^n} = \frac{\mathbf{v}\mathbf{u}^{\mathrm{T}}}{\mathbf{u}^{\mathrm{T}}\mathbf{v}}\,. \tag{6.19}$$

Then, given H1, the asymptotic behavior of the vector of mean values for the aggregated system is given by

$$\lim_{n\to\infty} \frac{E(\mathbf{Y}_n)}{\lambda^n} = \mathbf{l}^{\mathrm{T}}\mathbf{U}\mathbf{X}_0\mathbf{r}\,.$$

Now we have:

Proposition 7. *Given* H1, *then the asymptotic behavior of the vector of expected values for the original and auxiliary system is given by*

$$\lim_{n\to\infty} \frac{E(\mathbf{X}'_n)}{\lambda^n} = \frac{1}{\lambda}(\mathbf{l}^{\mathrm{T}}\mathbf{U}\mathbf{X}_0)\mathbf{M}\mathbf{V}\mathbf{r}\,,$$

$$\lim_{n\to\infty} \frac{E(\mathbf{X}_n)}{\lambda_k^n} = \frac{1}{\lambda}(\mathbf{l}^{\mathrm{T}}\mathbf{U}\mathbf{X}_0)\mathbf{M}\mathbf{V}\mathbf{r} + \mathbf{o}(\gamma^k)\,.$$

Proof. Direct consequence of (6.19), Lemma 3 and Proposition 3. □

6.2.2. *Second order moments*

As a tool for the subsequent developments we will characterize the asymptotic behavior of the second order moments of a general g.w.p. following the approach of Ref. 8. For this we will use the Kronecker matrix product,[9] which is defined for two matrices $\mathbf{A} = [a_{ij}] \in \mathbb{C}^{m\times n}$ and $\mathbf{B} = [b_{ij}] \in \mathbb{C}^{r\times s}$ as the matrix of size $mr \times ns$ given by

$$\mathbf{A}\otimes\mathbf{B} := \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{21}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}, \tag{6.20}$$

i.e. $\mathbf{A}\otimes\mathbf{B}$ is a matrix with mn blocks in which the block in position (i,j) has the form $a_{ij}\mathbf{B}$. A notable property of the Kronecker product that we will frequently use in the sequel is

$$(\mathbf{A}_1\otimes\mathbf{A}_2\otimes\cdots\otimes\mathbf{A}_m)(\mathbf{B}_1\otimes\mathbf{B}_2\otimes\cdots\otimes\mathbf{B}_m) = (\mathbf{A}_1\mathbf{B}_1)\otimes(\mathbf{A}_2\mathbf{B}_2)\otimes\cdots\otimes(\mathbf{A}_m\mathbf{B}_m)\,. \tag{6.21}$$

Moreover, for $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{m\times n}$ we define

$$\mathbf{A} := (a_{11}, a_{12}, \dots, a_{1n}, a_{21}, a_{22}, \dots, a_{2n}, \dots, a_{m1}, a_{m2}, \dots, a_{mn})^{\mathrm{T}} \in \mathbb{R}^{mn\times 1}\,.$$

Now, let $\mathbf{Z}_0$, $\mathbf{Z}_1$, $\mathbf{Z}_2, \dots$ be a generic g.w.p. with m types and let $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{m\times m}$ be the matrix of expected values of the system, i.e. $a_{ij} = E(z^i_{n+1}\,|\,\mathbf{Z}_n = e^j)$. Let $\mathbf{V}^j = [V^j_{rl}] \in \mathbb{R}^{m\times m}$ be the covariance matrix for the offspring of an individual of type j, i.e. $V^j_{rl} = \mathrm{Cov}(z^r_{n+1}, z^l_{n+1}|\mathbf{Z}_n = \mathbf{e}^j)\,;\ j,r,l = 1,\dots,m$ and let $\mathbf{C}^j(n) = E(\mathbf{Z}_n\mathbf{Z}_n^{\mathrm{T}}\,|\,\mathbf{Z}_0 = \mathbf{e}^j) = [C^j_{rs}(n)]$ be the matrix of moments of second order at time n, given the initial condition $\mathbf{e}^j$, i.e. $C^j_{rs}(n) = E(z^r_n z^s_n\,|\,\mathbf{Z}_0 = \mathbf{e}^j)$. Let

$$\mathbf{D} := (\mathbf{V}^1\,|\,\mathbf{V}^2\,|\,\cdots\,|\,\mathbf{V}^m) \in \mathbb{R}^{m^2\times m}\,,$$

Then the moments of first and second order of our system can be described by vector $\begin{pmatrix} E(\mathbf{Z}_n) \\ \mathbf{C}^j(n) \end{pmatrix}$ of dimension $(m+m^2)\times 1$. Following Ref. 8, the evolution of this vector is given by

$$\begin{pmatrix} E(\mathbf{Z}_{n+1}) \\ \mathbf{C}^j(n+1) \end{pmatrix} = \mathbf{\Omega} \begin{pmatrix} E(\mathbf{Z}_n) \\ \mathbf{C}^j(n) \end{pmatrix}, \tag{6.22}$$

where

$$\mathbf{\Omega} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{D} & \mathbf{A}\otimes\mathbf{A} \end{pmatrix} \in \mathbb{R}^{(m+m^2)\times(m+m^2)}$$

with $\mathbf{0}$ the $m\times m^2$ null matrix.

The next result characterizes the asymptotic behavior of the moments of second order in the supercritical and subcritical cases in terms of the dominant spectral characteristics of $\mathbf{A}$. The critical case is not discussed, for $\lambda = 1$ would be a double eigenvalue for $\mathbf{\Omega}$ and several different situations may arise in terms of the characteristics of $\mathbf{A}$ and $\mathbf{D}$.

Lemma 5. *In the conditions above, let $\mathbf{A}$ have a simple and strictly dominant eigenvalue λ with associated right and left eigenvectors $\mathbf{r}$ and $\mathbf{l}$ respectively. Assume that the initial population is $\mathbf{Z}_0 = \mathbf{e}^j$. Then we have:*

(a) *Supercritical case. If $\lambda > 1$, then the spectral radius of $\mathbf{\Omega}$ is λ^2 and right and left associated eigenvectors are $\begin{pmatrix} \mathbf{0} \\ \mathbf{r}\otimes\mathbf{r} \end{pmatrix}$ and $\begin{pmatrix} \mathbf{h} \\ \mathbf{l}\otimes\mathbf{l} \end{pmatrix}$ respectively, where $\mathbf{h} = (\lambda^2\mathbf{I} - \mathbf{A}^{\mathrm{T}})^{-1}\mathbf{D}^{\mathrm{T}}(\mathbf{l}\otimes\mathbf{l})$. Moreover,*

$$\lim_{n\to\infty} \frac{\mathbf{C}^j(n)}{\lambda^{2n}} = \frac{(\mathbf{l}^{\mathrm{T}}\otimes\mathbf{l}^{\mathrm{T}})((\mathbf{e}^j\mathbf{e}^{jT}) + \mathbf{D}(\lambda^2\mathbf{I}-\mathbf{A})^{-1}\mathbf{e}^j)}{(\mathbf{l}^{\mathrm{T}}\mathbf{r})^2}(\mathbf{r}\otimes\mathbf{r})\,.$$

(b) *Subcritical case. If $\lambda < 1$, then the spectral radius of $\mathbf{\Omega}$ is λ and right and left associated eigenvectors are $\begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix}$ and $\begin{pmatrix} \mathbf{l} \\ \mathbf{0} \end{pmatrix}$ respectively, where $\mathbf{s} = (\lambda\mathbf{I} - \mathbf{A}\otimes\mathbf{A})^{-1}\mathbf{D}\mathbf{r}$. Besides*

$$\lim_{n\to\infty} \frac{\mathbf{C}^j(n)}{\lambda^n} = \frac{\mathbf{l}^{\mathrm{T}}\mathbf{e}^j}{\mathbf{l}^{\mathrm{T}}\mathbf{r}}(\lambda\mathbf{I} - \mathbf{A}\otimes\mathbf{A})^{-1}\mathbf{D}\mathbf{r}\,.$$

Proof. Let $\lambda_1,\dots,\lambda_n$ be the eigenvalues of $\mathbf{A}$ with $\lambda = |\lambda_1| > |\lambda_2| \ge \cdots \ge |\lambda_n|$. The eigenvalues of $\mathbf{A}\otimes\mathbf{A}$ are the product "all with all" of the eigenvalues of $\mathbf{A}$ (see Ref. 9). The eigenvalues of $\mathbf{\Omega}$ are then the set $\{\lambda_i,\ \lambda_j\lambda_l\,;\ i,j,l = 1,\dots,m\}$.

(a) If $\lambda > 1$, then it is obvious that $\rho(\mathbf{\Omega}) = \lambda^2$ (being a simple and strictly dominant eigenvalue) and $\begin{pmatrix} \mathbf{0} \\ \mathbf{r}\otimes\mathbf{r} \end{pmatrix}$ is an associated right eigenvector. Since λ^2 cannot be an eigenvalue of $\mathbf{A}$, then $(\lambda^2\mathbf{I} - \mathbf{A})$ and $(\lambda^2\mathbf{I} - \mathbf{A}^{\mathrm{T}})$ are invertible matrices.

If we define $\mathbf{W}_1 = (\lambda^2\mathbf{I} - \mathbf{A})^{-1}$, then $\mathbf{h}^{\mathrm{T}} = (\mathbf{l}^{\mathrm{T}} \otimes \mathbf{l}^{\mathrm{T}})\mathbf{D}\mathbf{W}_1$. Furthermore, it is straightforward to check that $\mathbf{W}_1\mathbf{A} + \mathbf{I} = \lambda^2\mathbf{W}_1$. Now

$$\begin{aligned}(\mathbf{h}^{\mathrm{T}} \,|\, \mathbf{l}^{\mathrm{T}} \otimes \mathbf{l}^{\mathrm{T}})\boldsymbol{\Omega} &= ((\mathbf{l}^{\mathrm{T}} \otimes \mathbf{l}^{\mathrm{T}})\mathbf{D}[\mathbf{W}_1\mathbf{A} + \mathbf{I}] \,|\, \mathbf{l}^{\mathrm{T}}\mathbf{A} \otimes \mathbf{l}^{\mathrm{T}}\mathbf{A}) \\ &= ((\mathbf{l}^{\mathrm{T}} \otimes \mathbf{l}^{\mathrm{T}})\mathbf{D}\lambda^2\mathbf{W}_1 \,|\, \lambda\mathbf{l}^{\mathrm{T}} \otimes \lambda\mathbf{l}^{\mathrm{T}}) = \lambda^2(\mathbf{h}^{\mathrm{T}} \,|\, \mathbf{l}^{\mathrm{T}} \otimes \mathbf{l}^{\mathrm{T}})\end{aligned}$$

and therefore $\begin{pmatrix} \mathbf{h} \\ \mathbf{l} \otimes \mathbf{l} \end{pmatrix}$ is a left eigenvector of $\boldsymbol{\Omega}$ associated to λ^2. Now, since λ^2 is a simple and strictly dominant eigenvalue for $\boldsymbol{\Omega}$, we have

$$\begin{aligned}\lim_{n\to\infty} \frac{\begin{pmatrix} E(\mathbf{Z}_n) \\ \mathbf{C}^j(n) \end{pmatrix}}{\lambda^{2n}} &= \lim_{n\to\infty} \frac{\boldsymbol{\Omega}^n}{\lambda^{2n}} \begin{pmatrix} E(\mathbf{Z}_0) \\ \mathbf{C}^j(0) \end{pmatrix} \\ &= \frac{\begin{pmatrix} \mathbf{0} \\ \mathbf{r} \otimes \mathbf{r} \end{pmatrix} (\mathbf{h}^{\mathrm{T}} \,|\, \mathbf{l}^{\mathrm{T}} \otimes \mathbf{l}^{\mathrm{T}}) \begin{pmatrix} E(\mathbf{Z}_0) \\ \mathbf{C}^j(0) \end{pmatrix}}{(\mathbf{h}^{\mathrm{T}} \,|\, \mathbf{l}^{\mathrm{T}} \otimes \mathbf{l}^{\mathrm{T}}) \begin{pmatrix} \mathbf{0} \\ \mathbf{r} \otimes \mathbf{r} \end{pmatrix}} \\ &= \frac{\left(\mathbf{h}^{\mathrm{T}}\mathbf{Z}_0 + (\mathbf{l}^{\mathrm{T}} \otimes \mathbf{l}^{\mathrm{T}})\mathbf{C}^j(0)\right)}{(\mathbf{l}^{\mathrm{T}}\mathbf{r}) \otimes (\mathbf{l}^{\mathrm{T}}\mathbf{r})} \begin{pmatrix} \mathbf{0} \\ \mathbf{r} \otimes \mathbf{r} \end{pmatrix} \\ &= \frac{\left(\mathbf{l}^{\mathrm{T}} \otimes \mathbf{l}^{\mathrm{T}}\right)\left(\mathbf{D}\mathbf{W}_1\mathbf{e}^j + (\mathbf{e}^j\mathbf{e}^{jT})\right)}{(\mathbf{l}^{\mathrm{T}}\mathbf{r})^2} \begin{pmatrix} \mathbf{0} \\ \mathbf{r} \otimes \mathbf{r} \end{pmatrix}\end{aligned}$$

as we wanted to show. We have used $E(\mathbf{Z}_0) = \mathbf{e}^j$ and $\mathbf{C}_0^j = \mathbf{e}^j\mathbf{e}^{jT}$.

(b) If $\lambda < 1$, then it is obvious that $\rho(\boldsymbol{\Omega}) = \lambda$ and besides it is simple and strictly dominant, for if we had $|\lambda_i\lambda_j| = |\lambda|$ for some i and j, then we would have $|\lambda| = |\lambda_i\lambda_j| \leq |\lambda|^2$ which is impossible. Therefore matrix $(\lambda\mathbf{I} - \mathbf{A} \otimes \mathbf{A})$ is invertible. Checking that $\begin{pmatrix} \mathbf{l} \\ \mathbf{0} \end{pmatrix}$ is a left eigenvector of $\boldsymbol{\Omega}$ associated to λ is straightforward. If we define $\mathbf{W}_2 = (\lambda\mathbf{I} - \mathbf{A} \otimes \mathbf{A})^{-1}$ we have $\mathbf{s} = \mathbf{W}_2\mathbf{D}\mathbf{r}$ and, besides, $(\mathbf{A} \otimes \mathbf{A})\mathbf{W}_2 + \mathbf{I} = \lambda\mathbf{W}_2$. Now

$$\boldsymbol{\Omega}\begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\mathbf{r} \\ [\mathbf{I} + (\mathbf{A} \otimes \mathbf{A})\mathbf{W}_2]\mathbf{D}\mathbf{r} \end{pmatrix} = \begin{pmatrix} \lambda\mathbf{r} \\ \lambda\mathbf{W}_2\mathbf{D}\mathbf{r} \end{pmatrix} = \lambda\begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix}.$$

Also

$$\lim_{n\to\infty} \frac{\begin{pmatrix} E(\mathbf{Z}_n) \\ \mathbf{C}^j(n) \end{pmatrix}}{\lambda^n} = \frac{\begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix} (\mathbf{l}^{\mathrm{T}} \,|\, \mathbf{0}^{\mathrm{T}}) \begin{pmatrix} E(\mathbf{Z}_0) \\ \mathbf{C}^j(0) \end{pmatrix}}{(\mathbf{l}^{\mathrm{T}} \,|\, \mathbf{0}^{\mathrm{T}}) \begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix}} = \frac{\mathbf{l}^{\mathrm{T}}\mathbf{e}^j}{\mathbf{l}^{\mathrm{T}}\mathbf{r}} \begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix}$$

as we wanted to show. $\square$

Note that, in the supercritical case, the moments of second order of the population vector have an asymptotic rate of increase of λ^2, meanwhile the asymptotic structure of this vector is characterized by $\mathbf{r}\otimes\mathbf{r}$. In the subcritical case, the asymptotic rate of increase of the moments of second order is λ and the asymptotic structure of this vector is defined by $\mathbf{s} = (\lambda\mathbf{I} - \mathbf{A}\otimes\mathbf{A})^{-1}\mathbf{D}\mathbf{r}$.

Coming back to our aggregated system we have the following. For each $j = 1,\dots,q$, let $\bar{\mathbf{C}}^j(n) = [\bar{C}^j_{rs}(n)] = E(\mathbf{Y}_n\mathbf{Y}_n^{\mathrm{T}}\,|\,\mathbf{Y}_0 = \mathbf{e}^j) \in \mathbb{R}^{q\times q}$ be the matrix of second order moments for the aggregated system at time n, given an initial individual in group j and let

$$\bar{\mathbf{D}} = (\bar{\mathbf{V}}^1\,|\,\bar{\mathbf{V}}^2\,|\,\cdots\,|\,\bar{\mathbf{V}}^q) \in \mathbb{R}^{q^2\times q}\,.$$

Assume that initially there is only one individual in group j, i.e. $\mathbf{Y}_0 = \mathbf{e}^j$. Then, from H1 and Lemma 5 we have that the asymptotic behavior of the second order moments of this system is characterized in the following way:

(a) *Supercritical case.* If $\lambda > 1$, then

$$\lim_{n\to\infty}\frac{(\bar{\mathbf{C}}^j(n))}{\lambda^{2n}} = \frac{(\mathbf{l}^{\mathrm{T}}\otimes\mathbf{l}^{\mathrm{T}})((\mathbf{e}^j\mathbf{e}^{jT}) + \bar{\mathbf{D}}(\lambda^2\mathbf{I} - \bar{\mathbf{M}})^{-1}\mathbf{e}^j)}{(\mathbf{l}^{\mathrm{T}}\mathbf{r})^2}(\mathbf{r}\otimes\mathbf{r})\,.$$

(b) *Subcritical case.* If $\lambda < 1$, then

$$\lim_{n\to\infty}\frac{(\bar{\mathbf{C}}^j(n))}{\lambda^{n}} = \frac{\mathbf{l}^{\mathrm{T}}\mathbf{e}^j}{\mathbf{l}^{\mathrm{T}}\mathbf{r}}(\lambda\mathbf{I} - \bar{\mathbf{M}}\otimes\bar{\mathbf{M}})^{-1}\bar{\mathbf{D}}\mathbf{r}\,.$$

Let us consider the $N\times N$ matrices $\mathbf{C}^{jl}_k(n)$ and $\mathbf{C}'^j(n)$ of second order moments for the original and auxiliary system at time n. Note that, according to (5.9), the matrices corresponding to the auxiliary system are independent of the subgroup of the father.

$$\mathbf{C}^{jl}_k(n) = E(\mathbf{X}_n\mathbf{X}_n^{\mathrm{T}}\,|\,\mathbf{X}_0 = \mathbf{e}^{jl})\,;\;\; \mathbf{C}'^j(n) = E(\mathbf{X}'_n\mathbf{X}'^T_n\,|\,\mathbf{X}_0 = \mathbf{e}^{jl})\,,$$

$$j = 1,\dots,q\,;\; l = 1,\dots,N_j\,.$$

Besides, let

$$\mathbf{D}' = (\mathbf{V}'^1\,|\,\overset{(N_1)}{\cdots}\,|\,\mathbf{V}'^1\,|\,\mathbf{V}'^2\,|\,\overset{(N_2)}{\cdots}\,|\,\mathbf{V}'^2\,|\,\cdots\,|\,\mathbf{V}'^q\,|\,\overset{(N_q)}{\cdots}\,|\,\mathbf{V}'^q) \in \mathbb{R}^{N^2\times N}\,. \tag{6.23}$$

Then, the asymptotic behavior of the moments of second order for the original system and the auxiliary system is characterized by the following result:

Proposition 8. *Assume* $\mathbf{X}_0 = \mathbf{e}^{jl}$, $j = 1,\dots,q$, $l = 1,\dots,N_q$. *Given H1 we have*:

(a) *Supercritical case. If* $\lambda > 1$ *then*

$$\lim_{n\to\infty}\frac{(\mathbf{C}'^j(n))}{\lambda^{2n}}$$

$$= \frac{1}{\lambda^2}\frac{(\mathbf{l}^{\mathrm{T}}\mathbf{U}\otimes\mathbf{l}^{\mathrm{T}}\mathbf{U})((\mathbf{e}^{jl}\mathbf{e}^{jlT}) + \mathbf{D}'(\lambda^2\mathbf{I} - \mathbf{M}\bar{\mathbf{P}})^{-1}\mathbf{e}^{jl})}{(\mathbf{l}^{\mathrm{T}}\mathbf{r})^2}(\mathbf{MVr}\otimes\mathbf{MVr})\,,$$

$$\lim_{n\to\infty} \frac{(\mathbf{C}_k^{jl}(n))}{\lambda_k^{2n}}$$

$$= \frac{1}{\lambda^2} \frac{(\mathbf{l}^{\mathrm{T}}\mathbf{U} \otimes \mathbf{l}^{\mathrm{T}}\mathbf{U})((\mathbf{e}^{jl}\mathbf{e}^{jlT}) + \mathbf{D}'(\lambda^2\mathbf{I} - \mathbf{M}\bar{\mathbf{P}})^{-1}\mathbf{e}^{jl})}{(\mathbf{l}^{\mathrm{T}}\mathbf{r})^2} (\mathbf{MVr} \otimes \mathbf{MVr}) + \mathbf{o}(\gamma^k)\,.$$

(b) *Subcritical case. If* $\lambda < 1$, *then*

$$\lim_{n\to\infty} \frac{(\mathbf{C}'^j(n))}{\lambda^n} = \frac{1}{\lambda} \frac{\mathbf{l}^{\mathrm{T}}\mathbf{U}\mathbf{e}^{jl}}{\mathbf{l}^{\mathrm{T}}\mathbf{r}} (\lambda\mathbf{I} - \mathbf{M}\bar{\mathbf{P}} \otimes \mathbf{M}\bar{\mathbf{P}})^{-1}\mathbf{D}'\mathbf{MVr}\,,$$

$$\lim_{n\to\infty} \frac{(\mathbf{C}_k^{jl}(n))}{\lambda_k^n} = \frac{1}{\lambda} \frac{\mathbf{l}^{\mathrm{T}}\mathbf{U}\mathbf{e}^{jl}}{\mathbf{l}^{\mathrm{T}}\mathbf{r}} (\lambda\mathbf{I} - \mathbf{M}\bar{\mathbf{P}} \otimes \mathbf{M}\bar{\mathbf{P}})^{-1}\mathbf{D}'\mathbf{MVr} + \mathbf{o}(\gamma^k)\,.$$

Proof. Let us define the matrix $\mathbf{D}_k$ which is the analog, for the original system, of matrix $\mathbf{D}'$ in (6.23). The components of $\mathbf{G}_k$ and $\mathbf{G}'$ are polynomials and so are their second derivatives. Moreover, their coefficients are linear functions of the probabilities $p_k^{ij}(\alpha)$ and $p'^i(\alpha)$ respectively, and so it is straightforward to show, using Lemma 4, that $\mathbf{D}_k = \mathbf{D}' + \mathbf{o}(\gamma^k)$. Now the result follows directly from Lemma 5 and property (6.21) of the Kronecker product. □

6.3. *Subcritical case: mean population conditional on non-extinction*

It is known that for a subcritical positively regular g.w.p. $\mathbf{Z}_n$ there exists an asymptotic stationary distribution for the population conditional on non-extinction which, moreover, is independent of the (nonzero) initial population (see Ref. 14 for a complete discussion of the subcritical case). In other words, for each $\alpha \in \mathbb{Z}_+^m$ and each $z_0 \neq \mathbf{0}$, $\lim_{n\to\infty} pr(\mathbf{Z}_n = \alpha \mid \mathbf{Z}_n \neq \mathbf{0},\, \mathbf{Z}_0 = \mathbf{z}_0)$ exists and is independent of $\mathbf{z}_0$. Consider the generic g.w.p. under the hypothesis that $\mathbf{A}$ is primitive. Let us consider the dominant eigenvectors $\mathbf{r}$ and $\mathbf{l}$ of $\mathbf{A}$ normalized in the way $\|\mathbf{r}\|_1 = 1$, $\mathbf{l}^{\mathrm{T}}\mathbf{r} = 1$. Then we have that the limit

$$\xi := \lim_{n\to\infty} \frac{\mathbf{r}^{\mathrm{T}}(\mathbf{1} - \mathbf{q}(n))}{\lambda^n}$$

exists and, under very general conditions that are trivially satisfied if the p.g.f. has only a finite number of summands, is strictly positive. Moreover, we have

$$\lim_{n\to\infty} E(\mathbf{Z}_n \mid \mathbf{Z}_n \neq \mathbf{0},\, \mathbf{Z}_0 = \mathbf{z}_0) = \frac{1}{\xi}\mathbf{r} \tag{6.24}$$

for any $\mathbf{z}_0 \neq \mathbf{0}$.

We know that, for k large enough, the subcriticality of the original system is equivalent to that of the aggregated system. Let us now address the study of the relationship between the expected value of the population conditional on non-extinction for the original system and the aggregated system. We introduce the following hypothesis on the aggregated system.

H2. Let the aggregated system be positively regular (i.e. $\bar{\mathbf{M}}$ primitive) and subcritical (i.e. $\lambda < 1$) and let $\mathbf{M}$ be row allowable. For the dominant eigenelements of $\bar{\mathbf{M}}$ we will use the same notation and normalization conditions as in H1.

Observe that, according to Proposition 4, H2 guarantees that the auxiliary system is also subcritical and positively regular, and so is the original system for large enough k. Then we know that the following three limits:

$$\bar{\xi} := \lim_{n\to\infty} \frac{\mathbf{r}^{\mathrm{T}}(\mathbf{1} - \mathbf{q}(n))}{\lambda^n}, \quad \xi' := \lim_{n\to\infty} \frac{(\mathbf{MVr})^{\mathrm{T}}(\mathbf{1} - \mathbf{q}'(n))}{\lambda^n \|\mathbf{MVr}\|_1},$$

$$\xi_k := \lim_{n\to\infty} \frac{\mathbf{r}_k^{\mathrm{T}}(\mathbf{1} - \mathbf{q}_k(n))}{(\lambda_k)^n \|\mathbf{r}_k\|_1},$$

exist and are positive.

From H2 we have, for the aggregated system, that given any nonzero initial condition $\mathbf{y}_0$

$$\lim_{n\to\infty} E(\mathbf{Y}_n \,|\, \mathbf{Y}_n \neq \mathbf{0},\, \mathbf{Y}_0 = \mathbf{y}_0) = \frac{1}{\bar{\xi}}\mathbf{r}\,. \tag{6.25}$$

The following proposition, which constitutes our main result, characterizes the asymptotic behavior of the mean population conditional on non-extinction for the original system:

Proposition 9. *Given* H2 *holds, then*:

(a) $\xi' = \frac{\lambda}{\|\mathbf{MVr}\|_1}\bar{\xi}$.
(b) $\lim_{k\to\infty} \xi_k = \xi' = \frac{\lambda}{\|\mathbf{MVr}\|_1}\bar{\xi}$.
(c) *For the auxiliary and the original systems we have, respectively,*

$$\lim_{n\to\infty} E(\mathbf{X}'_n \,|\, \mathbf{X}_n \neq \mathbf{0},\, \mathbf{X}_0 = \mathbf{x}_0) = \frac{1}{\lambda\bar{\xi}}\mathbf{MVr}$$

$$\lim_{k\to\infty}\lim_{n\to\infty} E(\mathbf{X}_n \,|\, \mathbf{X}_n \neq \mathbf{0},\, \mathbf{X}_0 = \mathbf{x}_0) = \frac{1}{\lambda\bar{\xi}}\mathbf{MVr}\,,$$

for any initial population $\mathbf{x}_0 \neq \mathbf{0}$.

Note that, for large enough k, we can approximate the asymptotic behavior of the mean population vector for the original system conditional on non-extinction in terms of the parameters λ, $\bar{\xi}$ and $\mathbf{r}$ corresponding to the reduced model.

Proof. In the first place, let us point out a preliminary result regarding any subcritical positively regular g.w.p. $\mathbf{Z}_n$ with m types, matrix of expected values $\mathbf{A}$ and $\lambda = \rho(\mathbf{A})$. Since for each $j = 1, \ldots, m$ we have

$$E(\mathbf{Z}_n \,|\, \mathbf{Z}_n \neq \mathbf{0},\, \mathbf{Z}_0 = \mathbf{e}^j) = \frac{E(\mathbf{Z}_n \,|\, \mathbf{Z}_0 = \mathbf{e}^j)}{1 - pr(\mathbf{Z}_n = \mathbf{0} \,|\, \mathbf{Z}_0 = \mathbf{e}^j)}$$

$$= \frac{E(\mathbf{Z}_n \,|\, \mathbf{Z}_0 = \mathbf{e}^j)}{1 - q^j(n)} = \frac{1}{1 - q^j(n)}\mathbf{A}^n\mathbf{e}^j\,,$$

then, passing to the limit $n \to \infty$ and using (6.24) and (6.19) we obtain

$$\xi = \frac{1}{l^j} \lim_{n\to\infty} \frac{1 - q^j(n)}{\lambda^n} \quad \text{for all } j = 1, \ldots, q\,. \tag{6.26}$$

(a) It follows from Lemma 3 that $\frac{\|\mathbf{MVr}\|_1}{\lambda}\mathbf{U}^t\mathbf{l}$ is the dominant left eigenvector of the auxiliary system meeting the normalization requirements of H1, and so its (ij)th component is $\frac{\|\mathbf{MVr}\|_1}{\lambda} l^i$. Now, applying (6.26) to the auxiliary system we obtain $\xi' = \frac{\lambda}{\|\mathbf{MVr}\|_1} \frac{1}{l^i} \lim_{n\to\infty} \frac{1-q'^i(n)}{\lambda^n}$ for all $i = 1, \ldots, q$, and the result follows using the fact that $q'^i(n) = q^i(n)$ for all n.

(b) Analogously, from Proposition 3 we obtain that the dominant left eigenvector of the original system meeting the normalization requirements of H1 has the form $\frac{\|\mathbf{MVr}\|_1}{\lambda}\mathbf{U}^t\mathbf{l} + \mathbf{o}(\gamma^k)$, so that its (ij)th component can be written as $\frac{\|\mathbf{MVr}\|_1}{\lambda} l^i + o(\gamma^k)$ and then $\xi_k = \frac{\lambda}{\|\mathbf{MVr}\|_1 l^i + o(\gamma^k)} \lim_{n\to\infty} \frac{1-q_k^{ij}(n)}{(\lambda_k)^n}$ for all $i = 1, \ldots, q;, j = 1, \ldots, N_i$. From here it follows that:

$$\lim_{k\to\infty} \xi_k = \frac{\lambda}{\|\mathbf{MVr}\|_1 l^i} \lim_{k\to\infty} \lim_{n\to\infty} \frac{1 - q_k^{ij}(n)}{(\lambda_k)^n} = \frac{\lambda}{\|\mathbf{MVr}\|_1 l^i} \lim_{n\to\infty} \lim_{k\to\infty} \frac{1 - q_k^{ij}(n)}{(\lambda_k)^n}$$

$$= \frac{\lambda}{\|\mathbf{MVr}\|_1 l^i} \lim_{n\to\infty} \frac{1 - q'^i(n)}{\lambda^n} = \xi'\,,$$

where we have used that $\lim_{k\to\infty} \lambda_k = \lambda$ (Proposition 3), $\lim_{k\to\infty} q_k^{ij}(n) = q'^i(n)$ (Proposition 5) and that the two limits $n \to \infty$ and $k \to \infty$ in $\frac{1-q_k^{ij}(n)}{(\lambda_k)^n}$ can be interchanged due to the uniform convergence guaranteed by Lemma A.1 (see the Appendix).

(c) Straightforward from (a), (b) and (6.24). □

6.4. *Numerical simulations*

The previous results show that the main parameters related to the dynamics of the original system can be approximated in terms of information corresponding to the aggregated model. Those results characterize the approximations in the limit $k \to \infty$. In order to show that the approximations are good for finite values of k and consequently our aggregation procedure is useful for the study of real biological populations, we have performed some numerical simulations. These simulations correspond to the age-structured multiregional model of Sec. 3 and its corresponding reduced model, and we have chosen two of the most relevant parameters, growth rate of the expected population size and probability of ultimate extinction, to carry out the study.

Regarding the error we make when approximating $\lambda_k \approx \lambda$, Proposition 3 suggests that it depends on both, $|\gamma_{q+1}|$ (the modulus of the subdominant eigenvalue of $\mathbf{P}$) and k (the separation between the time scales of the slow and fast processes). Accordingly, for each $h \in \{0.1, 0.2, \ldots, 0.9\}$ we have constructed 50 random examples of original systems such that $|\gamma_{q+1}| = h$. Then, in each example we have

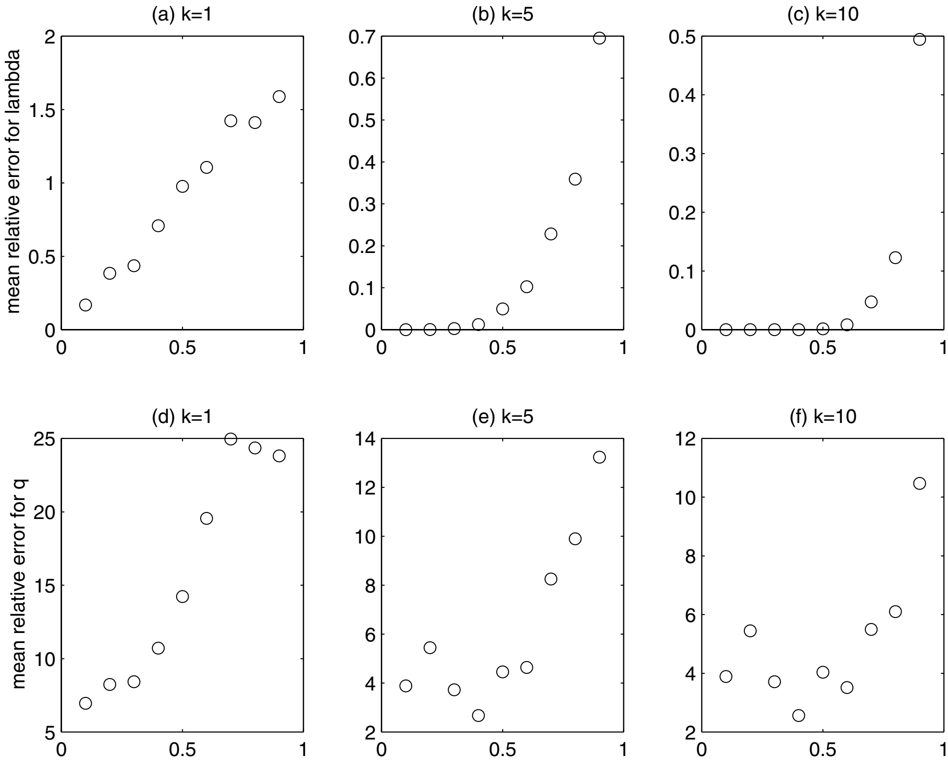

Fig. 1. Mean relative error obtained in the approximation of the growth rate and the probability of extinction. For each value of $|\gamma_{q+1}| = 0.1, 0.2, \ldots, 0.9$, we randomly generated 50 models. In each case we compared the quantities obtained on the aggregated system with those observed on the original system for different values of k.

obtained the corresponding aggregated system and compared λ_k and λ in one hand and $\mathbf{q}_k$ and $\bar{\mathbf{q}}$ in the other, for different values of k.

Figure 1(a) shows how the mean relative error generated in the approximation of the growth rate increases with $|\gamma_{q+1}|$ when k is fixed. Here, the maximum mean error obtained over the 450 generated models was about 1.5% corresponding to the case $k = 1$. Figures 1(b) and (c) show that these deviations decrease when k grows, becoming near to 0.5% for $k = 10$.

In the approximation of the extinction probability (see Figs. 1(d)–(f) the maximum mean error increases moderately taking values around 25% for $k = 1$ and falling down to 11% for $k = 10$.

The simulations carried out suggest a direct relationship between the accuracy of the approximations and both quantities: $|\gamma_{q+1}|$ and k. Furthermore, it is shown that, even for small values of k, the maximum mean error obtained is less than 10% in most of the cases except for some specific instances with high values of $|\gamma_{q+1}|$ which corresponds to the case in which the fast process converges very slowly. In summary, the aggregation method proposed seems appropriate for describing the dynamics of real biological populations.

Acknowledgment

This work has been partially supported by Proyecto de Investigación BFM2002-01615 (MCYT).

Appendix

Lemma A.1. *Under hypothesis* H2, *there exists* k_0 *such that the sequence* $\frac{\mathbf{1}-\mathbf{q}_k(n)}{(\lambda_k)^n}$ *converges to a certain vector* $\mathbf{b}$ *as* $n \to \infty$ *uniformly for* $k \geq k_0$.

Proof. Let us define $\mathbf{G}_k(n, \mathbf{s}) = \mathbf{G}_k \circ \mathbf{G}_k \circ \overset{(n)}{\cdots} \circ \mathbf{G}_k \circ \mathbf{G}_k(\mathbf{s})$ with components $G_k^{ir}(n, \mathbf{s}),\ i = 1, \ldots, q\ ;\ r = 1, \ldots, N_i$ and let

$$\begin{aligned} \mathbf{h}_k(n) &= (h_k^{11}(n), \ldots, h_k^{1N_1}(n), \ldots, h_k^{q1}(n), \ldots, h_k^{qN_q}(n)) \\ &:= \mathbf{G}_k(n, \mathbf{0}) - \mathbf{1} = \mathbf{q}_k(n) - \mathbf{1}\,. \end{aligned}$$

For each $i = 1, \ldots, q$, $r = 1, \ldots, N_i$ and each n and n' we have, by the general theory of g.w.p., $h_k^{ir}(n + n') = G_k^{ir}(n + n', \mathbf{0}) - 1 = G_k^{ir}(n, \mathbf{G}_k(n', \mathbf{0})) - 1$. Now, using Taylor's theorem for $G_k^{ir}(n, *)$ around $\mathbf{s} = \mathbf{1}$ and the fact that $G_k^{ir}(n, \mathbf{1}) = 1$ we obtain

$$h_k^{ir}(n + n') = \mathbf{h}_k(n')\mathrm{grad}G_k^{ir}(n, \mathbf{1}) + \frac{1}{2}\mathbf{h}_k(n')\mathbf{H}_k^{ir}(n, n')\mathbf{h}_k^{\mathrm{T}}(n')\,, \tag{A.1}$$

where $\mathbf{H}_k^{ir}(n, n')$ is the Hessian matrix of $G_k^{ir}(n, *)$ evaluated at a point of the segment that joins $\mathbf{1}$ and $\mathbf{G}_k(n', \mathbf{0}) = \mathbf{q}_k(n') \geq \mathbf{0}$ and $\mathrm{grad}G_k^{ir}(n, \mathbf{1})$ corresponds to the (ir)th column of $(\mathbf{MP}^k)^n$. Now we define the column vector $\mathbf{f}_k^{ir}(n, n') := \frac{1}{2}\mathbf{H}_k^{ir}(n, n')\mathbf{h}_k^{\mathrm{T}}(n')$ and the matrix $\mathbf{F}_k(n, n') \in \mathbb{R}^{N \times N}$ such that its (ir)th column is $\mathbf{f}_k^{ir}(n, n')\ ;\ i = 1, \ldots, q\ ;\ r = 1, \ldots, N_i$. Using (A.1) we can write $\mathbf{h}_k(n + n') = \mathbf{h}_k(n')[(\mathbf{MP}^k)^n + \mathbf{F}_k(n, n')]$. Next we will show that $\frac{\mathbf{h}_k(n)}{(\lambda_k)^n}$ is a uniform Cauchy sequence for k large enough and the lemma will be proved.

Note that, from hypothesis H2 and Proposition 4, there exists k_0 such that the original process is positive regular (i.e. $\mathbf{MP}^k$ be primitive) for $k \geq k_0$. Moreover, since $\lambda_k \overset{k\to\infty}{\to} \lambda < 1$ (Proposition 3), given any σ, $\lambda < \sigma < 1$, we can choose k_0 such that $\lambda_k < \sigma$ for $k \geq k_0$.

In the remaining of the proof, let $\| * \|$ denote both the l_∞ norm in $\mathbb{R}^N$ and its associated matrix norm in $\mathbb{R}^{N \times N}$. Now let n_0, n_1 and n_2 be arbitrary and $k \geq k_0$. Then

$$\begin{aligned} &\left\| \frac{\mathbf{h}_k(n_0 + n_1)}{(\lambda_k)^{n_0+n_1}} - \frac{\mathbf{h}_k(n_0 + n_2)}{(\lambda_k)^{n_0+n_2}} \right\| \\ &\leq \left\| \frac{\mathbf{h}_k(n_0)}{(\lambda_k)^{n_0}} \right\| \left(\left\| \left(\frac{\mathbf{MP}^k}{\lambda_k} \right)^{n_1} - \left(\frac{\mathbf{MP}^k}{\lambda_k} \right)^{n_2} \right\| \right. \\ &\quad \left. + (\lambda_k)^{n_0} \left\| \frac{\mathbf{F}_k(n_0, n_1)}{(\lambda_k)^{n_0+n_1}} - \frac{\mathbf{F}_k(n_0, n_2)}{(\lambda_k)^{n_0+n_2}} \right\| \right) \end{aligned}$$

$$\leq \left\| \frac{\mathbf{h}_k(n_0)}{(\lambda_k)^{n_0}} \right\| \left(\left\| \left(\frac{\mathbf{MP}^k}{\lambda_k} \right)^{n_1} - \left(\frac{\mathbf{MP}^k}{\lambda_k} \right)^{n_2} \right\| \right.$$

$$\left. + \sigma^{n_0} \left\| \frac{\mathbf{F}_k(n_0,n_1)}{(\lambda_k)^{n_0+n_1}} - \frac{\mathbf{F}_k(n_0,n_2)}{(\lambda_k)^{n_0+n_2}} \right\| \right) .$$

Now, $\mathbf{MP}^k$ is primitive and so $\|(\frac{\mathbf{MP}^k}{\lambda_k})^{n_1} - (\frac{\mathbf{MP}^k}{\lambda_k})^{n_2}\|$ tends to zero when n_1 and n_2 tend to infinity. Moreover, taking into account that Lemma A.2 (see below) guarantees that $\|\frac{\mathbf{h}_k(n)}{(\lambda_k)^n}\|$ and $\|\frac{\mathbf{F}_k(n,n')}{(\lambda_k)^{n+n'}}\|$ are bounded as functions of n and n', uniformly for k large enough, the desired result follows. □

Lemma A.2. *There exists k_0 such that* (a) $\|(\frac{\mathbf{MP}^k}{\lambda_k})^n\|$, (b) $\|\frac{\mathbf{C}_k(n)}{(\lambda_k)^n}\|$, (c) $\|\frac{\mathbf{h}_k(n)}{(\lambda_k)^n}\|$ *and* (d) $\|\frac{\mathbf{F}_k(n,n')}{(\lambda_k)^{n+n'}}\|$ *are bounded as functions of n and n', uniformly for $k \geq k_0$.*

Proof. We know from (6.22) that $\begin{pmatrix} E(\mathbf{X}_n) \\ \mathbf{C}_k(n) \end{pmatrix} = (\mathbf{\Omega}_k)^n \begin{pmatrix} \mathbf{X}_0 \\ (\mathbf{X}_0\mathbf{X}_0^{\mathrm{T}}) \end{pmatrix}$ where $\mathbf{\Omega}_k = \begin{pmatrix} \mathbf{MP}^k & \mathbf{0} \\ \mathbf{D}_k & \mathbf{MP}^k \otimes \mathbf{MP}^k \end{pmatrix}$ and $\mathbf{D}_k$ was defined in the proof of Proposition 8. By inspection of $\mathbf{\Omega}_k$ it follows that, in order to prove (a) and (b) it suffices to prove that $\|(\mathbf{\Omega}_k/\lambda_k)^n\|$ is bounded as a function of n uniformly for k large enough. We know that, for large enough k, $\lambda_k < 1$ is a simple and strictly dominant eigenvalue for $\mathbf{MP}^k$. So, from Lemma 5, $\rho(\mathbf{\Omega}_k) = \lambda_k$ is also simple and strictly dominant and has associated right and left eigenvectors $\mathbf{a}_k = \begin{pmatrix} \mathbf{r}_k \\ \mathbf{s}_k \end{pmatrix}$ and $\mathbf{b}_k = \begin{pmatrix} \mathbf{l}_k \\ \mathbf{0} \end{pmatrix}$ where $\mathbf{s}_k = (\lambda\mathbf{I} - \mathbf{MP}^k \otimes \mathbf{MP}^k)^{-1}\mathbf{D}_k\mathbf{r}_k$. From $\mathbf{D}_k = \mathbf{D}' + \mathbf{o}(\gamma^k)$ (proof of Proposition 8) and Proposition 3 we have that $\mathbf{a}_k$ and $\mathbf{b}_k$ converge when $k \to \infty$, and so the desired result will follow if we prove that $(\frac{\mathbf{\Omega}_k}{\lambda_k})^n \overset{n\to\infty}{\to} \mathbf{a}_k\mathbf{b}_k^{\mathrm{T}}$ uniformly for k large enough. For each k let us consider the eigenvalues of $\mathbf{MP}^k$ and $\mathbf{M\bar{P}}$ ordered by decreasing modulus be, respectively, $\lambda_k > |\lambda_{k,2}| \geq \cdots \geq |\lambda_{k,N}|$ and $\lambda > |\bar{\lambda}_2| \geq \cdots \geq |\bar{\lambda}_N|$. Using a Jordan canonical decomposition of $\mathbf{\Omega}_k$ we can write $(\frac{\mathbf{\Omega}_k}{\lambda_k})^n - \mathbf{a}_k\mathbf{b}_k^{\mathrm{T}} = (\frac{\mathbf{E}_k}{\lambda_k})^n$ where $\mathbf{E}_k$ is a matrix such that $\rho(\mathbf{E}_k) = \max\{|\lambda_{k,2}|, (\lambda_k)^2\}$. From the continuity of the eigenvalues on the entries of the matrix it follows $\lambda_k \overset{k\to\infty}{\to} \lambda$ and $\lambda_{k,2} \overset{k\to\infty}{\to} \bar{\lambda}_2$ and so $\rho(\frac{\mathbf{E}_k}{\lambda_k}) \overset{k\to\infty}{\to} \max\{\frac{\bar{\lambda}_2}{\lambda}, \lambda\} < 1$. Therefore, there exists $\sigma < 1$ such that $\rho(\frac{\mathbf{E}_k}{\lambda_k}) < \sigma$ for k large enough. Consequently $(\frac{\mathbf{E}_k}{\lambda_k})^n \overset{n\to\infty}{\to} \mathbf{0}$ uniformly for k big enough and the result follows.

(c) Let us recall that the partial derivatives of any order of a p.g.f. evaluated at any point $\mathbf{s} \geq \mathbf{0}$ are always non-negative. Moreover, if $\mathbf{G}(\mathbf{s})$ denotes the p.g.f. of the generic g.w.p. $\mathbf{Z}_n$, in Lemma 5, we have that for all j, r and l, and each $\mathbf{s}$ verifying $\mathbf{0} \leq \mathbf{s} \leq \mathbf{1}$ it follows that:

$$0 \leq \frac{\partial^2 G^j}{\partial s_r \partial s_l}(\mathbf{s}) \leq \frac{\partial^2 G^j}{\partial s_r \partial s_l}(\mathbf{1}) = C^j_{rl} - \delta_{rl} a_{rj} \leq C^j_{rl}\,, \qquad \text{(A.2)}$$

where δ_{rl} denotes the Kronecker delta.

From (A.2) it follows that $\mathbf{H}_k^{ir}(n,n') \geq \mathbf{0}$ for any i, r, n and n'. Besides, by definition, $\mathbf{h}_k(n) \leq \mathbf{0}$ for all n and so $\mathbf{h}_k(n')\mathbf{H}_k^{ir}(n,n')\mathbf{h}_k^{\mathrm{T}}(n') \geq \mathbf{0}$ for all n and n'. Now, from (A.1), $\mathbf{0} \leq -\mathbf{h}_k(n+n') \leq -\mathbf{h}_k(n')(\mathbf{MP}^k)^n$, so $|\mathbf{h}_k(n+n')| \leq |\mathbf{h}_k(n')|(\mathbf{MP}^k)^n$. Using that for the l_∞ norm $|\mathbf{A}| \leq |\mathbf{B}| \Rightarrow \|\mathbf{A}\| \leq \|\mathbf{B}\|$, we have $\|\mathbf{h}_k(n+n')\| \leq \|\mathbf{h}_k(n')\|\|(\mathbf{MP}^k)^n\| \leq 2\|(\mathbf{MP}^k)^n\|$. Making $n'=0$ we obtain $\|\frac{\mathbf{h}_k(n)}{(\lambda_k)^n}\| \leq 2\|(\frac{\mathbf{MP}^k}{\lambda_k})^n\|$. Now, using part (a), the boundedness of $\|\frac{\mathbf{h}_k(n)}{(\lambda_k)^n}\|$ as a function of n, uniformly for k large enough follows.
(d) From (A.2) applied to $G_k^{ir}(n,\mathbf{s})$ we obtain $\mathbf{H}_k^{ir}(n,n') \leq \mathbf{C}_k^{ir}(n)$ and therefore $\|\mathbf{H}_k^{ir}(n,n')\| \leq \|\mathbf{C}_k^{ir}(n)\|$. Now, for each $i=1,\ldots,q$ and $r=1,\ldots,N_i$,

$$\begin{aligned}\left\|\frac{\mathbf{f}_k^{ir}(n,n')}{(\lambda_k)^{n+n'}}\right\| &= \frac{1}{2}\left\|\frac{\mathbf{H}_k^{ir}(n,n')\mathbf{h}_k^{\mathrm{T}}(n')}{(\lambda_k)^{n+n'}}\right\| \\ &\leq \frac{1}{2}\left\|\frac{\mathbf{H}_k^{ir}(n,n')}{(\lambda_k)^n}\right\|\left\|\frac{\mathbf{h}_k^{\mathrm{T}}(n')}{(\lambda_k)^{n'}}\right\| \leq \frac{1}{2}\left\|\frac{\mathbf{C}_k^{ir}(n)}{(\lambda_k)^n}\right\|\left\|\frac{\mathbf{h}_k^{\mathrm{T}}(n')}{(\lambda_k)^{n'}}\right\|\end{aligned}$$

and now the boundness of $\|\frac{\mathbf{F}_k(n,n')}{(\lambda_k)^{n+n'}}\|$ as a function of n and n' uniformly for k large enough follows using (b) and (c). $\square$

References

1. P. Auger, *Dynamics and Thermodynamics in Hierarchically Organized Systems, Applications in Physics, Biology and Economics* (Pergamon Press, 1989).
2. P. Auger and J. C. Poggiale, *Emergence of population growth models: Fast migration and slow growth*, *Journal of Theoretical Biology* **182** (1996) 99–108.
3. J. D. Biggins, H. Cohn and O. Nerman, *Multi-type branching in varying environment*, *Stoc. Proc. Appl.* **83** (1999) 357–400.
4. A. Blasco, L. Sanz and R. Bravo de la Parra, *Approximate reduction of multiregional birth–death models with fast migration*, *Math. Comput. Model.* **36** (2002) 47–65.
5. G. L. Block and L. J. S. Allen, *Population extinction and quasi-stationary behavior in stochastic density-dependent structured models*, *Bull. Math. Biol.* **62** (2000) 199–228.
6. R. Bravo de la Parra, E. Sánchez, O. Arino and P. Auger, *A discrete model with density dependent fast migration*, *Math. Biosci.* **157** (1999) 91–109.
7. H. Caswell, M. Fujiwara and S. Brault, *Declining survival probability threatens the North Atlantic right whale*, *Proc. Natl. Acad. Sci. USA* **96** (1999) 3308–3313.
8. H. Caswell, *Matrix Population Models. Construction, Analysis, and Interpretation* (Sinauer Associates, 2001).
9. M. Fiedler, *Special Matrices and their Applications in Numerical Mathematics* (Martinus Nijhoff Publishers, 1986).
10. K. E. Holsinger, *Demography and extinction in small populations*, in *Genetics, Demography and the Viability of Fragmented Populations*, eds. A. G. Young and G. M. Clarke (Cambridge Univ. Press, 2000), pp. 55–74.
11. R. Horn and C. Johnson, *Matrix Analysis* (Cambridge Univ. Press, 1985).
12. Y. Iwasa, S. A. Levin and V. Andreasen, *Agregation in model ecosystems II. Approximate aggregation*, *IMA J. Math. Appl. Medicine Biol.* **6** (1989) 1–23.
13. P. Jagers, *Branching Processes with Biological Applications* (John Wiley & Sons, 1975).
14. A. Joffe and F. Spitzer, *On multi-type branching processes with $\rho \leq 1$*, *J. Math. Anal. Appl.* **19** (1967) 409–430.

15. S. Legendre, J. Clobert, A. P. Moller and G. Sorci, *Demographic stochasticity and social mating system in the process of extinction of small populations*: *The case of passerines introduced to New Zealand*, *The Am. Natural.* **153** (1999) 449–463.
16. C. J. Mode, *Multitype Branching Processes. Theory and Applications* (Elsevier, 1971).
17. L. Sanz and R. Bravo de la Parra, *Variables aggregation in a time discrete linear model*, *Math. Biosci.* **157** (1999) 111–146.
18. L. Sanz and R. Bravo de la Parra, *Time scales in stochastic multiregional models*, *Nonlinear Anal.*: *Real World Appl.* **1** (2000) 89–122.
19. L. Sanz and R. Bravo de la Parra, *Time scales in a non autonomous linear discrete model*, *Math. Models Methods in Appl. Sci.* **11** (2001) 1203–1235.
20. S. Tuljapurkar, *Population Dynamics in Variable Environments* (Springer-Verlag, 1990).
21. G. Wang, W. D. Edge and J. O. Wolff, *Demographic uncertainty in ecological risk assessments*, *Ecol. Model.* **136** (2001) 95–102.