Approximating the distribution of population size in stochastic multiregional matrix models with fast migration

BY JUAN ANTONIO ALONSO AND LUIS SANZ*

Departamento de Matemáticas, E.T.S.I. Industriales, Universidad Politécnica de Madrid, José Gutiérrez Abascal, 2, 28006 Madrid, Spain

In this work we deal with a multiregional model in discrete time for an age-structured population which lives in an environment that changes randomly with time and is distributed in different spatial patches. In addition, and as is often the case in applications, we assume that migration is fast with respect to demography. Using approximate aggregation techniques we make use of the existence of different time scales in the model and reduce the dimension of the system obtaining a stochastic Leslie model in which the variables are the total population in each age class. Literature shows that, under reasonable conditions, the distribution of population size in matrix models with environmental stochasticity is asymptotically lognormal, and is characterized by two parameters, stochastic growth rate (s.g.r.) and scaled logarithmic variance (s.l.v.), that, in most practical cases, cannot be computed exactly. We show that the s.g.r. and the s.l.v. of the original multiregional model can be approximated by those corresponding to the reduced stochastic Leslie model, therefore simplifying its analysis. Moreover, we illustrate the usefulness of the reduction procedure by presenting some practical cases in which, although the explicit computation of the s.g.r. and the s.l.v. of the original multiregional model is not feasible, we can calculate its analogues for the reduced model.

Keywords: approximate aggregation; environmental stochasticity; time scales; multiregional models; lognormal distribution

1. Introduction

The so-called ‘approximate aggregation methods’ are mathematical techniques in which appropriate approximations are introduced in order to transform the system under consideration into a reduced or ‘aggregated’ system with a lesser number of variables called ‘global variables’. These techniques are usually applied in systems governed by processes with different time scales, and the approximation usually involves freezing the slow process and letting the fast process reach equilibrium. Once this is done one can build a reduced or ‘aggregated’ system. The problem then is to justify that one can extract important

*Author for correspondence (lsanz@etsii.upm.es).

One contribution of 17 to a Theme Issue ‘From biological and clinical experiments to mathematical models’.
information about the behaviour of the original system through the study of the simpler reduced one. The reduction in the dimension of the system makes these techniques useful to study some of the complex models that we find in nature in which there are different time scales involved.

Approximate aggregation techniques have been widely studied in the context of time continuous systems with different-time scales for both linear and density-dependent models (see Auger et al. (2008) for a review and a list of references). We will focus our attention on the discrete time case, which has been thoroughly explored in linear, nonlinear and nonautonomous deterministic contexts (see among others Marvá et al. (2008); Sanz & Bravo de la Parra (1999)).

Nowadays, a great part of the research in ecological modelling is devoted to models which incorporate stochasticity as a way of taking into account different factors which either have an intrinsic random nature or are too complex to be taken into account in a deterministic way. The literature contemplates two sources of stochasticity: demographic (Caswell 2001) and environmental. The reduction of models that incorporate demographic stochasticity has been addressed in Sanz et al. (2003).

In this work we contemplate linear discrete models that incorporate ‘environmental stochasticity’, i.e. the randomness introduced when we consider random fluctuations in the environment and, consequently, in the vital rates which affect the population (see Caswell (2001) and Tuljapurkar & Caswell (1997, ch. 3) for an introduction; Tuljapurkar (1990) for a full discussion; and Cohen et al. (1983) for an application to actual fish populations). These models are analogous to the deterministic ones but in this case the matrix of vital rates in each projection interval is selected within a given set of matrices according to a certain (possibly time-varying) probability distribution. Given certain hypotheses on the pattern of temporal variation and the vital rates in each environment, the distribution of total population size is asymptotically lognormal, with an expected value and a variance dependent on two parameters: the stochastic growth rate (s.g.r.), which is the stochastic analogue of the logarithm of the dominant eigenvalue for deterministic systems, and the scaled logarithmic variance (s.l.v.), which characterizes the asymptotic deviation of the population from its mean value. In most practical situations it is not possible to compute the s.g.r. and the s.l.v. analytically and one must resort to computer simulations.

The reduction of discrete time systems for populations subjected to environmental stochasticity is carried out in Sanz & Bravo de la Parra (2000), and both this reference and Sanz & Bravo de la Parra (2007) provide certain results that relate the behaviour of the original and the aggregated system. In particular, in the case in which there are a finite number of environments, Sanz & Bravo de la Parra (2007) relates the s.g.r. of the original and the aggregated system.

The purpose of this work is to extend the abovementioned results in two directions. In the first place, and placing ourselves in a context in which the number of environmental states can be infinite, we prove that under reasonable assumptions the s.g.r. and the s.l.v. of the original system can be approximated by those of the aggregated system when the separation of time scales between demography and migration is large enough. This allows us to approximate the population distribution of the original system in terms of the s.g.r. and the s.l.v. of the reduced system.
In the second place, we show how, in some particular cases, the aggregation procedure renders a reduced system for which the s.g.r. and the s.l.v. can be computed exactly. This allows one to carry out a qualitative study of the behaviour of these models as functions of their parameters that is not possible when dealing directly with the original system. Moreover, both the reduction procedure and the results that relate the original and the aggregated systems are valid in a wider context than that of multiregional models, including any stage-structured model for a population living in a multipatch system.

The structure of the paper is as follows: §2 briefly introduces the basic form of the matrix models that consider environmental stochasticity and the main results that will be used in the rest of the paper, concerning the distribution of population size when the environmental variation is a homogeneous Markov chain. Section 3 is devoted to presenting a stochastic multiregional model in which population is structured by age and patch and migration among the different patches is fast with respect to demography. By defining the global variables as the total population in each age class, we obtain a reduced stochastic Leslie model. The model, together with the aggregation procedure, is a particular case of a technique which was first presented by the authors in Sanz & Bravo de la Parra (2000) and which is valid, for example, for any stage-structured population living in a multipatch environment.

Section 4 presents the abovementioned relationships between the original system and the reduced system. The proof of the most technical results is deferred to an appendix. In §5, where ecologists mainly interested in applications may focus their attention, we show how in some cases a non-trivial multiregional system, for which the s.g.r. and s.l.v. cannot be calculated exactly, can be aggregated to a stochastic Leslie model for which that calculation is possible. We first examine a non-structured population living in a multipatch environment and then we deal with a population with juveniles and adults living in a two-patch environment.

2. Matrix models with environmental stochasticity

This section presents the basic form of the matrix models that consider environmental stochasticity. We assume then that the population lives in an ambient in which there are different environmental states that we suppose indexed with $\zeta \in \mathcal{I}$, where $\mathcal{I}$ is a (possibly infinite, numerable or denumerable) set of indexes. The vital rates corresponding to each one of these environments are given by the non-negative matrices $A_\zeta \in \mathbb{R}^{N \times N}$, $\zeta \in \mathcal{I}$ in such a way that, for each $\zeta$, $A_\zeta$ represents the vital rates of the population in environment $\zeta$ and $\mathcal{A} = \{A_\zeta, \zeta \in \mathcal{I}\}$ is the set of different vital rates.

The environmental variation is characterized by a sequence of random variables $\tau_n$, $n = 0, 1, 2, \ldots$ defined on a certain probability space $(\Omega, \mathcal{F}, p)$ and with state space $\mathcal{I}$. For each realization $\omega \in \Omega$ of the process, the population is subjected to environmental conditions $\tau_{n+1}(\omega)$ between times $n$ and $n + 1$.

Thus, the model reads:

$$Z_{n+1} = A_{\tau_{n+1}} Z_n,$$

where for each $n = 0, 1, \ldots, Z_n = (z_1^n, \ldots, z_N^n)^T$ is a vector random variable in $\mathbb{R}^N$ which represents the population vector at time $n$. Throughout we will assume
that $Z_0$ is a fixed (non-random) non-zero vector $z_0 \geq 0$. We will also consider the structure of the population vector

$$W_n := \frac{Z_n}{\|Z_n\|},$$

(2.2)

where $\|*\|$ denotes (unless specifically stated) the 1-norm in $\mathbb{R}^N$, i.e. $\|Z_n\| = |z_1^n| + |z_2^n| + \cdots + |z_N^n|$. Let $\mathcal{W} = \{w \in \mathbb{R}^N : w \geq 0, \|w\| = 1\}$ be the set of possible population structures (column vectors with non-negative components and unitary in the 1-norm) and let $\mathcal{B}$ be the $\sigma$-algebra of Borel sets of $\mathcal{W}$.

Following a great part of the literature on the field, we will deal exclusively with the case in which the environmental variation $\tau_n$ is a homogeneous Markov chain (which includes as a particular case the i.i.d. setting). Let us define the $n$-step transition probabilities

$$p^{(n)}_{\tau C} = \mathbb{P}\{\tau_n \in C \mid \tau_1 = \xi\}, \quad C \subset \mathcal{I}, \quad \xi \in \mathcal{I}.$$ (2.3)

Then it is immediate to check that the bivariate process $(D_n)_{n=1}^\infty = ((\tau_n, W_n))_{n=1}^\infty$ is a homogeneous Markov chain with state space $\mathcal{I} \times \mathcal{W}$. Following Cohen (1977) we define the transition probabilities of this chain as

$$G_n(\xi, w, C \times B) := \mathbb{P}\{\tau_n \in C, W_n \in B \mid \tau_1 = \xi, W_0 = w\},$$

for $\xi \in \mathcal{I}, w \in \mathcal{W}, C \subset \mathcal{I}, B \in \mathcal{B}$.

Let us recall some features regarding the asymptotic behaviour of general matrix models with environmental stochasticity like equation (2.1). In order to do so, we introduce the following concept.

**Definition 2.1.** A set $A = \{A_{\xi}, \xi \in \mathcal{I}\}$ of $N \times N$ non-negative matrices is said to be ergodic if there exists a positive integer $g$ such that any product of $g$ matrices (with repetitions allowed) drawn from $A$ is a positive matrix (i.e. its components are all positive) and moreover there exist constants $\alpha, \beta > 0$ such that for all $\xi \in \mathcal{I}$

$$\max(A_{\xi}) \leq \beta; \min^+(A_{\xi}) \geq \alpha,$$

where $\max(A_{\xi})$ is the maximum of the elements of $A_{\xi}$ and $\min^+(A_{\xi})$ is the minimum of the positive elements of $A_{\xi}$.

In particular, any finite set of $N \times N$ non-negative matrices which share a common primitive incidence matrix is an ergodic set. More generally, if the condition $C \geq A_{\xi} \geq B, \xi \in \mathcal{I}$ with $B$ and $C$ primitive matrices holds, then $A$ is an ergodic set. Note that if the set $A$ is ergodic and $z_0 \geq 0, z_0 \neq 0$ then $Z_n \neq 0$ for all $n$ with probability 1 and so the population structure (2.2) is well defined.

In what follows we will assume that the vital rates and the pattern of environmental variation for system (2.1) meet the following hypotheses.

(H1) The chain $\tau_n$ has a unique stationary probability distribution $\mathbb{P}_\pi$ and moreover

$$|p^{(n)}_{\tau C} - \mathbb{P}_\pi(C)| \leq D_1 \rho^n \quad \text{for } C \subset \mathcal{I}, \quad \xi \in \mathcal{I},$$

(2.4)

where $0 \leq D_1 < \infty, 0 \leq \rho < 1$ are constants.

(H2) The set $A$ of matrices of vital rates is ergodic.
If there is a finite number of environmental conditions, i.e. if \( \tau_n \) takes only a finite number of values then hypothesis (H1) is met if the chain is irreducible and aperiodic, i.e. if its transition probabilities matrix is primitive (Doob 1953).

Under hypotheses (H1) and (H2), from Cohen (1977) we have that system (2.1) is stochastically strongly ergodic in the sense that the \( n \)-step transition probabilities \( G_n \) for system (2.1) converge to a stationary probability distribution \( F \) uniformly in the initial state of the chain. Specifically,

**Theorem 2.2.** (Cohen 1977, theorem 3.) Let us assume that system (2.1) meets hypotheses (H1) and (H2). Then there exists a stationary probability distribution \( F \) for \( \mathbf{D}_n \) such that, given \( \varepsilon > 0 \), \( C \subset \mathcal{I} \) and \( B \in \mathcal{B} \), \( \exists n_0 \in \mathbb{N} \) such that \( \forall n \geq n_0, \xi \in \mathcal{I}, \ w \in \mathcal{W} \)

\[
|G_n(\xi, w, C \times B) - F(C \times B)| \leq \varepsilon.
\]

Now let us turn to the study of the total population size \( \|Z_n\| \). In this regard we have the following theorem.

**Theorem 2.3.** (Tuljapurkar & Orzack 1980, essentially lemma 4.) Under hypotheses (H1) and (H2),

(i) We can define the s.g.r. of system (2.1) through

\[
a := \lim_{n \to \infty} \frac{\log \|Z_n\|}{n},
\]

with probability 1. Moreover, \( a \) is finite, and is independent of the initial probabilities of the chain and of the initial (non-zero) population vector \( z_0 \geq 0 \). Parameter \( a \) can also be computed through

\[
a = \lim_{n \to \infty} \frac{\mathbb{E} \log \|Z_n\|}{n} = \mathbb{E}_F \left[ \log \left( \mathbf{A}_{\tau_2} \mathbf{W}_1 \right) \right] = \mathbb{E}_F \left[ \log \left( \frac{\left( \mathbf{A}_{\tau_2} \mathbf{W}_1 \right)_1}{\left( \mathbf{W}_1 \right)_1} \right) \right],
\] (2.5)

where the averages with the \( F \) subscript are taken with respect to the stationary distribution \( F \) and assuming that \( z_0 \) is positive.

(ii) We can also define the s.l.v. of the system as

\[
\sigma^2 := \lim_{n \to \infty} \frac{1}{n} \mathbb{V}[\log \|Z_n\|],
\]

where \( \mathbb{V} \) denotes variance and where \( \sigma^2 \) is independent of the initial probabilities of the chain and of the initial (non-zero) population vector \( z_0 \geq 0 \). Moreover, \( \sigma^2 \) is finite and

\[
\sigma^2 = c_1 + 2 \sum_{r=2}^{\infty} c_r,
\]

where

\[
c_r := \mathbb{E}_F \left[ \left( \log \left( \frac{\left( \mathbf{A}_{\tau+1} \mathbf{W}_r \right)_1}{\left( \mathbf{W}_r \right)_1} - a \right) \right) \left( \log \left( \frac{\left( \mathbf{A}_{\tau_2} \mathbf{W}_1 \right)_1}{\left( \mathbf{W}_1 \right)_1} - a \right) \right) \right] ; \ r = 1, 2, 3, \ldots
\] (2.6)

*Phil. Trans. R. Soc. A* (2009)
(iii) If $\sigma^2 > 0$ the population size is asymptotically lognormal in the sense that

$$\log \|Z_n\| - na \sim \mathcal{L} \to \mathcal{N}(0,1)$$

where $\mathcal{N}(0,1)$ denotes a normal distribution of zero mean and unit variance and $\mathcal{L}$ denotes convergence in distribution.

Proof. Essentially this result is lemma 4 in Tuljapurkar & Orzack (1980) but in this reference the authors impose an additional condition (expression (11') in Tuljapurkar & Orzack (1980) or, equivalently, condition (3.13) in Furstenberg & Kesten (1960)) that, we will show, is not necessary given (H1) and (H2). Indeed, let us define variables

$$\xi_g := \log(T_{g,11}^{j}), \quad \xi_n := \log \left[ \frac{(T_n)_{11}^{j}}{(T_{n-1})_{11}^{j}} \right] \text{ for each } n > g; \quad \text{with } T_n = A_{\tau_n} \ldots A_{\tau_1}.$$

Lemma 4 of Tuljapurkar & Orzack (1980) hinges on theorem 3 in Furstenberg & Kesten (1960), in which the authors apply the central limit theorem to the variables $\xi_n$ (which in that reference are not necessarily bounded) and they need an additional condition to guarantee that the moments of $\xi_n$ are finite. However, using (H2) it is straightforward to check that there exist constants $C$ and $D$ such that,

$$0 < C \leq \frac{(T_n)_{ij}}{(T_{n-1})_{pq}} \leq D; \quad 1 \leq i, j, p, q \leq N, \quad n > g,$$

with $C$ and $D$ depending only on the constants $g, \alpha, \beta$ corresponding to the ergodic set $\mathcal{A}$. Therefore, $|\xi_n|$ is bounded for each $n$ and therefore its moments are finite. The rest of the proof is identical to that of Tuljapurkar & Orzack (1980).}

The convergence to the lognormal distribution for the population size has also been established under similar but weaker hypotheses in Heyde & Cohen (1985).

Note that if $a > 0$ then $\|Z_n\|$ grows exponentially for every realization and if $a < 0$ then the population goes extinct with probability 1.

It is important to stress that, in most practical cases, the exact computation of $a$ and $\sigma^2$ is not feasible due to the fact that in general one cannot obtain an analytical expression for the stationary probability distribution $F$ that intervenes in equations (2.5) and (2.6). In §5 we make use of the aggregation procedure to be able to obtain exact results for certain multiregional models with fast migration.

### 3. Reduction of a stochastic multiregional model with fast migration

In this section we introduce the model and the reduction procedure for a certain kind of stochastic multiregional models in which migration is fast with respect to demography. The procedure is a particular case of that introduced in Sanz & Bravo de la Parra (2000) to reduce general models which incorporate environmental stochasticity and in which there are two time scales.

Multiregional models consider the dynamics of a possibly age-structured population distributed among different spatial patches among which they can migrate. These models have been used for the study of human
populations (Rogers 1995) and in ecology (Caswell 2001). The usual approach has been deterministic, but the stochastic setting has also been used (Cohen 1982).

The above references do not explicitly consider the existence of different time scales in the multiregional system. The fact that in many situations migration and demography take place with different time scales (usually migration is fast with respect to demography although the contrary may also happen; see Liaw (1980)) has been exploited in several works in order to reduce the complexity of the multiregional model. Some examples in a linear deterministic setting are that of Sanz & Bravo de la Parra (1999) (autonomous case) and Sanz & Bravo de la Parra (2001) (non-autonomous case), and in the nonlinear case we refer the reader to Sanz et al. (2008). The use of multiregional models with different time scales to study specific instances of fish populations can be found in Charles et al. (2000).

We think of a population living in a habitat in which there are different environmental conditions, which we suppose indexed by an index \( \zeta \in \mathcal{I} \) that influences the vital rates. The population is structured in \( q \) age classes (corresponding to groups) and spread out in different spatial patches (subgroups) among which they may migrate. We assume that individuals in each age class \( i \) may migrate among \( r \) spatial patches. Therefore, the total number of subgroups is \( N := qr \) and the composition of the population is then given by vector \( X_n = (x_{n}^{11}, \ldots, x_{n}^{1r}, \ldots, x_{n}^{q1}, \ldots, x_{n}^{qr})^T \in \mathbb{R}^N \), where \( x_{n}^{ij} \) is the number of individuals in age class \( i \) living in patch \( j \) at time \( n \) and \( T \) denotes transposition.

Demography and migration are responsible for the transference of individuals among the different stages, and we suppose that migration is a fast process in comparison with demography. Moreover, we choose as the time step \( \Delta_n = (n, n + 1) \) for the model, the duration of each age class.

For each \( i \) and each environment \( \zeta \in \mathcal{I} \), migration for individuals of age \( i \) is modelled by a matrix \( P_i(\zeta) \in \mathbb{R}^{r \times r} \) that, since migration is a conservative process for the total number of individuals, is column stochastic. Besides, we suppose that \( P_i(\zeta) \) is primitive for each \( i \) and each \( \zeta \). This is the case, for example, if the fast process in each environment verifies: (i) transition from any patch to any other, in a sufficient number of steps, is allowed and (ii) individuals of at least one patch are allowed to stay in that patch. Finally, let \( P_\zeta = \text{diag}(P_1(\zeta), \ldots, P_q(\zeta)) \) be the matrix that models migration for the population and let the eigenvalues of \( P_\zeta \) (i.e. the union of the eigenvalues of the different \( P_i(\zeta) \)) ordered by decreasing modulus be

\[
1 = \lambda_1(\zeta) = \cdots = \lambda_q(\zeta) > |\lambda_{q+1}(\zeta)| \geq \cdots \geq |\lambda_N(\zeta)|; \quad \zeta \in \mathcal{I}.
\]

Now we assume that there exists \( \delta < 1 \) such that

\[
\forall \zeta \in \mathcal{I}, \quad |\lambda_{q+1}(\zeta)| \leq \delta. \quad (3.1)
\]

The demographic process in each environment \( \zeta \) is defined by the following generalized Leslie matrix

\[
M_\zeta = \begin{bmatrix}
F_1(\zeta) & F_2(\zeta) & \cdots & F_{q-1}(\zeta) & F_q(\zeta) \\
S_1(\zeta) & 0 & \cdots & 0 & 0 \\
0 & S_2(\zeta) & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & S_{q-1}(\zeta) & 0
\end{bmatrix},
\]

(3.2)

Phil. Trans. R. Soc. A (2009)
where

\[
F_i(\zeta) := \text{diag}(F^1_i(\zeta), \ldots, F^r_i(\zeta)), \quad i = 1, \ldots, q;
\]

\[
S_i(\zeta) := \text{diag}(S^1_i(\zeta), \ldots, S^r_i(\zeta)), \quad i = 1, \ldots, q - 1,
\]

and the coefficients \(F_j^i(\zeta)\) and \(S_j^i(\zeta)\) denote, respectively, the fertility and survival rates for individuals of age \(i\) in patch \(j\) in environment \(\zeta\).

In order to approximate the effect of migration over the time step of the model, which is much longer than its corresponding projection interval, we assume that if the population is subjected to environment \(\zeta\) during \(\Delta t\) matrix \(P_\sigma\) operates a number \(k\) of times, where \(k\) (which we assume to be an integer) can be interpreted as the ratio between the projection intervals corresponding to demography and migration. Thus, the set of vital rates for our system in the different environments is

\[
A^{(k)} := \{M_\zeta P_\zeta^k; \; \zeta \in \mathcal{I}\}. \tag{3.3}
\]

The pattern of environmental variation is defined by a sequence of random variables \(\tau_n, n = 1, 2, \ldots\) which take values in the set of environmental states \(\mathcal{I}\). Therefore, the proposed model, which we will refer to as the ‘original system’, consists in the following system of \(N\) random difference equations:

\[
X_{n+1} = M_{\tau_{n+1}} P_{\tau_{n+1}}^k X_n, \tag{3.4}
\]

with the initial condition \(X_0 = x_0\) with probability 1, where \(x_0 \geq 0\). We will denote the variables of the original system as \(X_n^{(k)}\) whenever we want to stress that they depend on the \(k\) chosen.

Now, following Sanz & Bravo de la Parra (2000), we make use of approximate aggregation to reduce the original system (3.4), consisting of \(N\) variables (microvariables) associated with the different subgroups, by an aggregated system of \(q\) variables (global variables), each of them associated with one group.

Let \(v_i(\zeta)\) be the right Perron eigenvector of \(P_i(\zeta)\), i.e. the eigenvector uniquely defined by the conditions \(P_i(\zeta)v_i(\zeta) = v_i(\zeta), v_i(\zeta) > 0, 1^T v_i(\zeta) = 1\) where \(1 = (1, \ldots, 1)^T\). \(v_i(\zeta)\) would characterize the equilibrium structure of migration for individuals in age class \(i\) if the population were constantly subjected to environment \(\zeta\).

In this way, the matrix that characterizes the conditions for migration in environment \(\zeta\) for group \(i\) is

\[
\bar{P}_i(\zeta) := \lim_{k \to \infty} P_{\tau_{n+1}}^k(\zeta) = v_i(\zeta) 1^T > 0, \quad \zeta \in \mathcal{I}, \quad i = 1, \ldots, q, \tag{3.5}
\]

and, for the total population, we have matrix

\[
\bar{P}_\zeta := \text{diag}(\bar{P}_1(\zeta), \bar{P}_2(\zeta), \ldots, \bar{P}_q(\zeta)).
\]

Note that, due to equation (3.1), we have that

\[
\sup_{\zeta \in \mathcal{I}} \|P^k(\zeta) - \bar{P}(\zeta)\| \to 0. \tag{3.6}
\]

Now we define matrices

\[
U := \text{diag}(1^T_1, 1^T_2, \ldots, 1^T_q); \quad V_\zeta := \text{diag}(v_1(\zeta), v_2(\zeta), \ldots, v_q(\zeta)), \quad \zeta \in \mathcal{I},
\]
and introduce the ‘auxiliary system’ as the stochastic model defined by
\[
\hat{X}_{n+1} = M_{\tau_{n+1}} \hat{P}_{\tau_{n+1}} \hat{X}_n,
\]
with the initial condition \( \hat{X}_0 = x_0 \), and let
\[
A_{aux} := \{ M_\zeta \hat{P}_\zeta ; \ \zeta \in \mathcal{I} \}
\]
be the set of matrices of vital rates for this system. Note that the auxiliary system can be interpreted as the result of letting migration reach equilibrium in the original system.

For each age class \( i \) we define a global variable \( y^i_n \) corresponding to the total population of the auxiliary system with age \( i \), i.e.
\[
y^i_n = \hat{x}^i_1 + \hat{x}^i_2 + \cdots + \hat{x}^i_r, \quad i = 1, \ldots, q,
\]
and then we define the vector \( Y_n \) of global variables
\[
Y_n = (y^1_n, \ldots, y^q_n)^T := U \hat{X}_n \in \mathbb{R}^q.
\]

From equation (3.5) it follows that \( \hat{P}_\zeta = V_\zeta U \) for all \( \zeta \in \mathcal{I} \). Now we multiply both sides of equation (3.7) by \( U \) and obtain
\[
U \hat{X}_{n+1} = U M_{\tau_{n+1}} \hat{P}_{\tau_{n+1}} \hat{X}_n = U M_{\tau_{n+1}} V_{\tau_{n+1}} U \hat{X}_n,
\]
which can be written in terms of the global variables exclusively, i.e. we have an aggregated system defined by
\[
Y_{n+1} = \tilde{M}_{\tau_{n+1}} Y_n, \quad (3.9)
\]
with the initial condition \( Y_0 = Ux_0 \) where, for each \( n \), \( \tilde{M}_{\tau_n} \) is given by
\[
\tilde{M}_{\tau_n} := U M_{\tau_n} V_{\tau_n}.
\]

The aggregated system (3.9) can be interpreted as a stochastic model in which the pattern of environmental variation coincides with that of the original system and in which the matrix of vital rates in each environment \( \zeta \) is
\[
\tilde{M}_\zeta := U M_\zeta V_\zeta \in \mathbb{R}^{q \times q}, \quad \zeta \in \mathcal{I},
\]
in such a way that the set of matrices for the different environmental conditions is
\[
A_{ag} := \{ \tilde{M}_\zeta ; \ \zeta \in \mathcal{I} \}.
\]

Note that, for each \( \zeta \), \( \tilde{M}_\zeta \) is a classical Leslie matrix given by
\[
\tilde{M}_\zeta = \begin{bmatrix}
f_1(\zeta) & f_2(\zeta) & \cdots & f_{q-1}(\zeta) & f_q(\zeta) \\
s_1(\zeta) & 0 & \cdots & 0 & 0 \\
0 & s_2(\zeta) & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & s_{q-1}(\zeta) & 0
\end{bmatrix},
\]

*Phil. Trans. R. Soc. A* (2009)
where the vital rates have the form

\[ f_i(\zeta) = \mathbf{1}^T \mathbf{F}_i(\zeta) \mathbf{v}_i(\zeta) = \sum_{j=1}^{r} v_i^j(\zeta) F_i^j(\zeta), \quad i = 1, \ldots, q; \quad \zeta \in \mathcal{I}, \]

\[ s_i(\zeta) = \mathbf{1}^T \mathbf{S}_i(\zeta) \mathbf{v}_i(\zeta) = \sum_{j=1}^{r} v_i^j(\zeta) S_i^j(\zeta), \quad i = 1, \ldots, q - 1; \quad \zeta \in \mathcal{I}, \]

i.e. each fertility rate \( f_i(\zeta) \) in the aggregated system is a weighted linear combination of the fertility rates in the general system corresponding to individuals of age class \( i \) in environment \( \zeta \), being the weights of the coefficients of the equilibrium spatial distribution for migration in environment \( \zeta \). Something analogous holds for the survival rates.

The original multiregional model has been transformed into a reduced system in which the spatial distribution has been averaged in a certain way and the population appears structured only by age.

As we mentioned above, both the reduction procedure we presented and the relationships between the original and the reduced systems are valid for a more general kind of model subjected to the effects of environmental stochasticity. Specifically, the results are valid for a whole class of models governed by different time scales presented in Sanz & Bravo de la Parra (2000), which in particular include the following settings:

(i) The population is divided into \( q \) groups (age classes in model (3.4)) attending to any characteristic of the life cycle, and each group \( i = 1, 2, \ldots, q \) is itself split into \( N_i \) subgroups (spatial patches in model (3.4)). Then, \( x_{ij}^n \) is the number of individuals in group \( i \) and subgroup \( j \) at time \( n \) and the population vector is \( \mathbf{X}_n = (x_{11}^n, \ldots, x_{1N_1}^n, \ldots, x_{q1}^n, \ldots, x_{qN_q}^n)^T \in \mathbb{R}^N \) where \( N := N_1 + \cdots + N_q \).

(ii) The projection interval of the model is that corresponding to the slow dynamics, on which we impose no special assumptions. Therefore, for any environmental condition \( \zeta \), the slow process will be modelled by a non-negative projection matrix \( \mathbf{M}_\zeta \in \mathbb{R}^{N \times N} \), which can be considered as divided into blocks \( \mathbf{M}_{ij}(\zeta), 1 \leq i, j \leq q \) of dimensions \( N_i \times N_j \) that characterize the rates of transference of individuals from the subgroups of group \( j \) to the subgroups of group \( i \) in environment \( \zeta \).

(iii) Regarding the fast process,

- 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, and
- for each \( i \) and \( \zeta \), matrix \( \mathbf{P}_i(\zeta) \in \mathbb{R}^{N_i \times N_i} \) is a column stochastic primitive matrix; in particular, the fast process is conservative of the total number of individuals in each group.

This setting allows one to model a population structured by any characteristic of its life cycle (stage structured models, see Caswell (2001) for comments and references) and living in a multipatch environment with migration among the different patches. A particular case of this is the case of spatially heterogeneous size-structured models.

We have to stress that both the aggregation procedure and the results obtained in the next section remain valid under the more general assumptions above.

Phil. Trans. R. Soc. A (2009)
4. Relationship between the distribution of population size in the original system and the aggregated system

Although as we have seen that the aggregation procedure is valid independently of the pattern of environmental variation for system (3.4), in order to study the relationships between the behaviour of the original and the aggregated system, we impose conditions under which the total population size is asymptotically lognormal. More specifically we will assume,

A1. The chain \( \tau_n \) verifies condition \((H1)\) in \( \S 2 \), i.e. it has a unique stationary probability distribution \( \mathbb{P}_\pi \) that verifies \((2.4)\).

Let us now investigate the relationships between the ergodicity of the sets \( \mathcal{A}(k) \) and \( \mathcal{A}_{ag} \). In order to do so let us introduce the following starting hypothesis:

A2. The non-zero entries of matrices \( M_\zeta \) and \( P_\zeta \) are bounded away from zero and infinity, i.e. there exist positive constants \( \epsilon, K, \epsilon' \) such that for all \( \zeta \in I \) we have

\[
\min^+ (M_\zeta) \geq \epsilon; \quad \max (M_\zeta) \leq K; \quad \min^+ (P_\zeta) \geq \epsilon'.
\] (4.1)

The next result shows that, under condition A2, the positive entries corresponding to the original and aggregated systems are bounded away from zero and infinity and, moreover, that given a very weak condition if \( \mathcal{A}_{ag} \) is an ergodic set then so are \( \mathcal{A}_{aux} \) and \( \mathcal{A}(k) \) for large enough \( k \). First we introduce the following definition: a non-negative matrix \( A \) is said to be row-allowable if it has at least a non-zero element in each one of its rows.

**Proposition 4.1.** (i) Assume A2 holds. Then there exist positive constants \( \alpha, \beta \) such that for all \( k \in \mathbb{N} \) and all \( \zeta \in I \) we have

\[
\min^+ (M_\zeta P_\zeta^k) \geq \alpha; \quad \min^+ (M_\zeta \bar{P}_\zeta) \geq \alpha; \quad \min^+ (\bar{M}_\zeta) \geq \alpha
\]

and

\[
\max (M_\zeta P_\zeta^k) \leq \beta; \quad \max (M_\zeta \bar{P}_\zeta) \leq \beta; \quad \max (\bar{M}_\zeta) \leq \beta.
\] (4.2)

i.e. positive entries of the matrices of the sets \( \mathcal{A}(k), \mathcal{A}_{aux} \) and \( \mathcal{A}_{ag} \) can be bounded away from zero and infinity and, moreover, the bounds can be chosen independent of \( k \).

(ii) If any product of \( g \) matrices drawn from \( \mathcal{A}_{ag} \) is positive and matrices \( M_\zeta \) are row-allowable, then any product of any \( g + 1 \) matrices drawn from \( \mathcal{A}_{aux} \) is positive and there exists \( k_0 \in \mathbb{N} \) such that for all \( k \geq k_0 \) the same thing happens for any product of \( g + 1 \) matrices drawn from \( \mathcal{A}(k) \).

**Proof.** (i) Since the entries of matrices \( P_\zeta(\zeta) \) are non-negative and its columns add up to 1, the entries of \( P_\zeta(\zeta) \) must be bounded from above for all \( i \) and \( \zeta \). In part (a) of proposition 3 in Sanz & Bravo de la Parra (2001) (where the role of the subindex \( n \) there is played by \( \zeta \) here) it is proved that if A2 holds then the normalized right eigenvectors \( v_i(\zeta) \) for matrices \( P_\zeta(\zeta) \) have entries bounded away from zero and infinity, i.e. for all \( \zeta \in I, 0 < \gamma \leq v_i(\zeta) \leq \gamma' \) for certain \( \gamma \) and \( \gamma' \).
Now, it is clear that if \( A \in \mathbb{R}^{t \times r} \) and \( B \in \mathbb{R}^{r \times l} \) are non-negative matrices and \( AB \neq 0 \), then \( \min^+(AB) \geq \min^+(A) \min^+(B) \) and \( \max(AB) \leq r \max(A) \max(B) \). Therefore, from \( A2 \) and the definition of \( \bar{P}_\xi \) and \( M_\xi \), it follows that there exist constants \( \alpha', \alpha'', \beta', \beta'' \), such that for all \( \xi \in \mathcal{I} \), \( \min^+(M_\xi \bar{P}_\xi) \geq \alpha' > 0 \); \( \min^+(\bar{M}_\xi) \geq \alpha'' > 0 \); \( \max(M_\xi \bar{P}_\xi) \leq \beta' \); \( \max(M_\xi) \leq \beta'' \). Now using this fact and the convergence of \( P_\xi^k \) to \( \bar{P}_\xi \) uniform in \( \xi \in \mathcal{I} \), it follows that there exist \( \alpha'' < \alpha' \) and \( \beta'' > \beta' \) such that for all \( k \in \mathbb{N} \), \( \min^+(M_\xi P_\xi^k) \geq \alpha'' > 0 \); \( \max(M_\xi P_\xi^k) \leq \beta'' \). The result follows taking \( \alpha = \min(\alpha'', \alpha') \) and \( \beta = \max(\beta'', \beta') \).

(ii) This result is part (a) of proposition 8 in Sanz & Bravo de la Parra (2007). The fact that here the set of matrices can be infinite does not introduce any changes in the proof.

As a consequence of the previous result we have,

**Corollary 4.2.** Assume \( A1 \) and \( A2 \) holds. If the set \( \mathcal{A}_{ag} \) is ergodic and the matrices \( M_\xi, \xi \in \mathcal{I} \) are row-allowable, then the set \( \mathcal{A}^{(k)} \) is ergodic for all \( k \geq k_0 \) (where \( k_0 \) is given by proposition 4.1). Moreover, the constants that bound above and below the positive entries of matrices of \( \mathcal{A}^{(k)} \) can be chosen independent of \( k \).

Therefore, if the aggregated system meets the sufficient conditions of theorems 2.2 and 2.3 for the existence of stochastic strong ergodicity, a s.g.r. and a s.l.v., and matrices \( M_\xi, \xi \in \mathcal{I} \) are row-allowable, then those sufficient conditions are also met by the auxiliary system and by the original system for \( k \geq k_0 \).

Let us introduce,

**A3.** The set \( \mathcal{A}_{ag} \) is ergodic and the matrices \( M_\xi, \xi \in \mathcal{I} \) are row-allowable.

Note that matrices (3.2) are row-allowable iff all \( S_j^i(\xi) \) are not zero and for every environment \( \xi \) and every patch \( j \) there is at least one age class \( i \) in which \( F_j^i(\xi) \neq 0 \). Moreover, if for all \( \xi \in \mathcal{I} \), \( A \leq M_\xi \leq B \) with \( A \) and \( B \) primitive matrices, then \( \mathcal{A}_{ag} \) is ergodic.

Therefore, under conditions \( A1 \)–\( A3 \), we can define the parameters

\[
\text{Aggregated system, } a = \lim_{n \to \infty} \frac{\mathbb{E} \log \|Y_n\|}{n} \quad \sigma^2 = \lim_{n \to \infty} \frac{1}{n} \mathbb{V}[\log \|Y_n\|],
\]

\[
\text{Original system, } a^{(k)} = \lim_{n \to \infty} \frac{\mathbb{E} \log \|X_n^{(k)}\|}{n} \quad \sigma^{(k)2} = \lim_{n \to \infty} \frac{1}{n} \mathbb{V}[\log \|X_n^{(k)}\|].
\]

We are now interested in finding a relationship between the parameters \( a^{(k)} \) and \( a \), on the one hand, and between \( \sigma^{(k)} \) and \( \sigma \), on the other hand. Specifically, we would like to be able to approximate \( a^{(k)} \) and \( \sigma^{(k)} \) through \( a \) and \( \sigma \), easier to compute when \( k \) is large. Theorem 9 in Sanz & Bravo de la Parra (2007) guarantees that we can carry out this approximation for the s.g.r. in the case in which the number of environmental states is finite, but the technique used in the proof does not allow one to establish a similar result for the s.l.v. Our main result guarantees, in this more general framework with possibly infinite environmental states, that both \( a^{(k)} \) and \( \sigma^{(k)} \) can be approximated by \( a \) and \( \sigma \).

Phil. Trans. R. Soc. A (2009)
Stochastic multiregional matrix models

Theorem 4.3. Let us assume conditions A1, A2 and A3. Then

$$\lim_{k \to \infty} a^{(k)} = a,$$

and

$$\lim_{k \to \infty} \sigma^{(k)} = \sigma,$$

and so the lognormal asymptotic distribution for the original system can be approximated through the parameters $a$ and $\sigma$ that correspond to the reduced system.

Proof. See the appendix. ■

As a consequence of this result we have that the lognormal asymptotic distribution for the original system can be approximated through that corresponding to the reduced system.

5. Applications: exact results for the reduced system

As we previously pointed out, in most ecological models the exact computation of the s.g.r. and the s.l.v. is not feasible, due to the fact that there is not an explicit expression for the stationary distribution of age structure. Therefore, one must resort to computer simulations to approximate those parameters (Tuljapurkar & Caswell 1997). The results of the last section allow one to carry out these simulations using the simpler aggregated system. Furthermore, in some cases our reduction procedure allows transformation of a non-trivial model into a simpler reduced one in which we can compute the s.g.r. and the s.l.v. exactly. Below we introduce two such cases.

(a) Non-structured population living in a multipatch environment

Let us consider the multiregional model of §3 in the case in which the population is not structured by age ($q = 1$), is distributed among $r$ spatial patches and there are a finite number $s$ of environmental conditions that vary according to a Markov chain $\tau_n$ with values in the set $\mathcal{I} = \{1, 2, \ldots, s\}$.

The population vector is $X_n = (x_1^n, \ldots, x_r^n) \in \mathbb{R}^r$, where $x_i^n$ is the population in patch $i$. Moreover we have that demography in environment $\zeta$ is modelled by $M_{\zeta} = \text{diag}(m_1(\zeta), m_2(\zeta), \ldots, m_r(\zeta)),$

where $m_i(\zeta), \ i = 1, \ldots, r; \ \zeta = 1, \ldots, s$ is the growth rate in patch $i$ and environment $\zeta$. Migration among the different patches in each environment $\zeta$ is modelled by a column-stochastic primitive matrix $P_{\zeta} \in \mathbb{R}^{r \times r}$. If we assume that all the growth rates are positive $m_i(\zeta) > 0, \ i = 1, \ldots, r; \ \zeta = 1, \ldots, s$, then condition A3 of §4 is met.

The Markov chain $\tau_n$ is defined by a certain vector $q \in \mathbb{R}^s$ of initial probabilities and a row-stochastic matrix $Q \in \mathbb{R}^{s \times s}$ of transition probabilities such that $q_{uv} := P\{\tau_{n+1} = v \mid \tau_n = u\}, \ 1 \leq u, v \leq s$. We assume that $\tau_n$ is irreducible and ergodic, i.e. that $Q$ is a primitive matrix, and so $\tau_n$ has a stationary probability distribution, which we denote $\pi = (\pi_1, \pi_2, \ldots, \pi_s)^T$, and condition A1 of §4 is met.
Let \( \mathbf{v}_\xi > 0 \) be the equilibrium distribution for migration in environment \( \xi \) normalized so that \( \| \mathbf{v}_\xi \| = 1 \). Then the reduced system (3.9) takes the form of a scalar stochastic system,

\[
y_{n+1} = \mathbf{m}_{\tau_{n+1}} y_n,
\]

where \( y_n = x_{n+1}^1 + \cdots + x_{n+1}^r \) is the total population and

\[
\mathbf{m}_\xi := \mathbf{1}^T \mathbf{M}_\xi \mathbf{v}_\xi = \sum_{i=1}^r v_i(\xi) m_i(\xi).
\]

Clearly the set \( \mathcal{A}_{\text{sc}} = \{ \mathbf{m}_1, \ldots, \mathbf{m}_s \} \) is ergodic and matrices \( \mathbf{M}_\xi \) are row-allowable, so both the reduced system (5.1) and the original system (for large enough \( k \)) meet the sufficient hypotheses for the existence of an asymptotic lognormal distribution. We can explicitly compute the s.g.r. and the s.l.v. that allowble, so both the reduced system (5.1) and the original system (for large enough \( k \)) meet the sufficient hypotheses for the existence of an asymptotic lognormal distribution. We can explicitly compute the s.g.r. and the s.l.v. that characterize the asymptotic distribution for system (5.1), the reason being that in this case the normalized population \( y_n/\|y_n\| \) is trivial (equal to 1 with probability 1), and so the stationary distribution of the chain \( (\tau_n, y_n/\|y_n\|) \) is simply that of \( \tau_n \).

In order to proceed we assume that \( \mathbf{Q} \) is diagonalizable so that \( \mathbf{Q} = \mathbf{RDR}^{-1} \), with \( \mathbf{D} = \text{diag}(1, \lambda_2, \ldots, \lambda_s) \) and \( \mathbf{R} = (\mathbf{1}, \mathbf{r}_2, \ldots, \mathbf{r}_s) \), where \( 1, \lambda_2, \ldots, \lambda_s \) are the eigenvalues of \( \mathbf{Q} \) and \( \mathbf{1}, \mathbf{r}_2, \ldots, \mathbf{r}_s \) corresponding right eigenvectors, \( \mathbf{Qr}_j = \lambda_j \mathbf{r}_j \), \( 2 \leq j \leq s \) and \( \mathbf{Q} \mathbf{1} = \mathbf{1} \).

Let us define the stochastic variables \( Z_{\tau_n} = \ln \mathbf{m}_{\tau_n} \). Then, according to theorem 2.3 the s.g.r. of system (5.1) is

\[
a = \mathbb{E}_\pi(Z_{\tau_n}) = \sum_{\xi=1}^s \pi_\xi \ln \mathbf{m}_\xi = \sum_{\xi=1}^s \pi_\xi \ln \left( \sum_{i=1}^r v_i(\xi) m_i(\xi) \right).
\]

Let us now define the variables (with zero mean with respect to the stationary distribution) \( \tilde{Z}_{\tau_n} := Z_{\tau_n} - a \), which take values \( z_\xi := \ln \mathbf{m}_\xi - a \), \( 1 \leq \xi \leq s \). Then the variables \( c_r \) defined in theorem 2.3 are

\[
c_1 = \mathbb{E}_\pi(\tilde{Z}_{\tau_n})^2, \quad c_r = \mathbb{E}_\pi(\tilde{Z}_{\tau_{n+r-1}} \tilde{Z}_{\tau_n}).
\]

The \( r \)-step transition probabilities are defined by matrix \( \mathbf{Q}^r \) so that \( q^{(r)}_{uv} = \mathbb{P}(\tau_{n+r} = v \mid \tau_n = u) \). A straightforward calculation shows that

\[
c_r = \mathbf{z}_\pi^T \mathbf{Q}^{r-1} \mathbf{z} = \mathbf{z}_\pi^T \mathbf{RD}^{r-1} \mathbf{R}^{-1} \mathbf{z},
\]

where the column vectors \( \mathbf{z} \) and \( \mathbf{z}_\pi \) are defined as

\[
\mathbf{z} = (z_1, z_2, \ldots, z_s)^T; \quad \mathbf{z}_\pi = (z_1 \pi_1, z_2 \pi_2, \ldots, z_s \pi_s)^T.
\]

Taking into account that \( \mathbf{R} = (\mathbf{1}, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_s) \) and the fact that \( \sum_{j=1}^s z_j \pi_j = 0 \) we have

\[
\mathbf{z}_\pi^T \mathbf{RD}^{r-1} = (0, \langle \mathbf{z}_\pi, \mathbf{v}_2 \rangle, \langle \mathbf{z}_\pi, \mathbf{v}_3 \rangle, \ldots, \langle \mathbf{z}_\pi, \mathbf{v}_s \rangle) \mathbf{D}^{r-1} = \langle 0, \lambda_2^{-1} \mathbf{z}_\pi, \lambda_3^{-1} \mathbf{z}_\pi, \ldots, \lambda_s^{-1} \mathbf{z}_\pi \rangle
\]

\[\text{Phil. Trans. R. Soc. A (2009)}\]
so that the s.l.v. $\sigma^2$ of the reduced system can be written as

$$\sigma^2 = c_1 + 2 \sum_{r=2}^{\infty} c_r = c_1 + 2 \sum_{r=2}^{\infty} z_r^T R D^{-1} R^{-1} z$$

$$= c_1 + 2 \left( \frac{\lambda_2}{1 - \lambda_2} (z_\pi, r_2), \frac{\lambda_3}{1 - \lambda_3} (z_\pi, r_3), \ldots, \frac{\lambda_s}{1 - \lambda_s} (z_\pi, r_s) \right) R^{-1} z. \quad (5.3)$$

Summing up, if the separation of time scales is large enough, the aggregation procedure allows one to approximate the s.g.r. and the s.l.v. of the original system, whose explicit computation is not feasible, through the explicit expressions (5.2) and (5.3).

For example, expression (5.2) makes it possible to study how changes in the parameters affect the long run growth ($a > 0$) or extinction ($a < 0$) of the population.

(b) A population with juveniles and adults living in a two-patch environment

(i) Model and reduction

Let us consider the setting of §3 in the case $q = 2$, and $r = 2$, i.e. the population has two age classes, juveniles and adults, and is living in a habitat with two patches. Then the population vector in the original system is $X_n = (x_{n11}, x_{n12}, x_{n21}, x_{n22})^T$ and

$$M_\xi = \begin{bmatrix} F_1(\xi) & 0 & F_1^2(\xi) & 0 \\ 0 & F_2(\xi) & 0 & F_2^2(\xi) \\ S_1(\xi) & 0 & 0 & 0 \\ 0 & S_2(\xi) & 0 & 0 \end{bmatrix};$$

$$P_\xi = \begin{bmatrix} 1 - p_1(\xi) & t_1(\xi) & 0 & 0 \\ p_1(\xi) & 1 - t_1(\xi) & 0 & 0 \\ 0 & 0 & 1 - p_2(\xi) & t_2(\xi) \\ 0 & 0 & p_2(\xi) & 1 - t_2(\xi) \end{bmatrix},$$

where $S^j(\xi)$ stands for the survival rate of juveniles in patch $j$, $p_i(\xi)$ denotes the migration rates in environment $\xi$ of individuals of age $i$ from patch 1 to patch 2 and $t_i(\xi)$ has an analogous meaning for migration from patch 2 to patch 1. If we assume that, for all $\xi \in \mathcal{I}$, $p_1(\xi)$, $p_2(\xi)$, $t_1(\xi)$ and $t_2(\xi)$ are different from 0 or 1 then matrices $P_1(\xi)$ and $P_2(\xi)$ are primitive for all $\xi \in \mathcal{I}$, and so we can reduce system (3.4).

Vectors $v_i(\xi)$ have the form $v_i(\xi) = (t_i(\xi)/(p_i(\xi) + t_i(\xi)), p_i(\xi)/(p_i(\xi) + t_i(\xi)))^T$, $i = 1, 2$, and therefore the reduced system (3.9) has the form

$$Y_{n+1} = \tilde{M}_{\tau_{n+1}} Y_n, \quad (5.4)$$
with \( Y_n = (y_1^n, y_2^n)^T = (x_1^n + x_2^n, x_1^n + x_2^n)^T \) and

\[
\bar{M}_\xi = \begin{bmatrix} f_1(\xi) & f_2(\xi) \\ s(\xi) & 0 \end{bmatrix},
\]

\[
s(\xi) = \frac{t_1(\xi) S^1(\xi) + p_1(\xi) S^2(\xi)}{p_1(\xi) + t_1(\xi)};
\]

\[
f_i(\xi) = \frac{t_i(\xi) F_{i1}(\xi) + p_i(\xi) F_{i2}(\xi)}{p_i(\xi) + t_i(\xi)}, \quad i = 1, 2; \quad \zeta \in \mathcal{I}.
\]

We proceed to analyse the reduced system (5.4) in three cases regarding the kind of environmental variation for which we can obtain exact results.

(ii) Case 1

Let us assume that the parameters that govern demography are independent of the environment, i.e. \( F_{ij}(\tau_n) = F_i^j, \ S^j(\zeta) = S^j, \) where \( F_i^j, S^j > 0, \ S^j < 1. \) Moreover, let

\[
p_i(\tau_n) = \eta_n p_i; \quad t_i(\tau_n) = \eta_n t_i,
\]

where \( p_i, t_i \in (0, 1) \) are non-random and \( \eta_n \) is a homogeneous Markov chain with values in \( [\varepsilon, 1] \), where \( 0 < \varepsilon < 1 \), that meets hypothesis A1 in §4 (with \( \eta_n \) playing the role of \( \tau_n \)). Note that hypothesis A2 is also met.

This model can be interpreted by saying that reproduction and survival are not affected by the environment and that migration rates depend on the environment through a common multiplicative random variable. A good/bad environment increases/decreases all migration rates by the same factor.

Then we have

\[
\bar{M}_\xi = \bar{M} = \begin{bmatrix} f_1 & f_2 \\ s & 0 \end{bmatrix},
\]

where

\[
s := \frac{t_1 S^1 + p_1 S^2}{p_1 + t_1}; \quad f_i := \frac{t_i F_{i1} + p_i F_{i2}}{p_i + t_i}, \quad i = 1, 2,
\]

i.e. the reduced system is the deterministic system \( Y_{n+1} = \bar{M} Y_n \). Since \( \bar{M} \) is a primitive matrix hypothesis A3 of §4 is met and so the results of that section hold.

Note that the s.g.r. and the s.l.v. of the reduced system are

\[
a = \log \frac{f_1 + \sqrt{(f_1)^2 + 4 f_2 s}}{2}; \quad \sigma^2 = 0.
\]

(iii) Case 2

Let us assume that migration is deterministic, i.e. \( p_i(\zeta) = p_i, \ t_i(\zeta) = t_i \) (with \( p_i, t_i \in (0, 1) \)) and that

\[
F_{ij}^i(\tau_n) = f_{ij}^i \eta_{\tau_n}; \quad S^j(\tau_n) = s^j \eta_{\tau_n}; \quad i = 1, 2,
\]
where we assume that \( f_i^j, s^j \) are non-random, positive and \( s^j < 1 \). \( \eta_{\tau_n} \) is a scalar homogeneous Markov chain, which we assume irreducible and ergodic that takes values in the set \( \{ \delta_1, \delta_2, \ldots, \delta_s \} \), where \( \delta_i \in (0, 1) \]. We denote its stationary probability distribution as \( \pi \).

Note that this case can be interpreted by saying that all fertility and survival rates depend on the environment through a common multiplicative random variable. A good/bad environment increases/decreases all vital rates by the same factor.

Then the aggregated system is \( \mathbf{Y}_n = \bar{\mathbf{M}}_{\tau_n} \mathbf{Y}_{n-1} \) where \( \bar{\mathbf{M}}_{\tau_n} = \eta_{\tau_n} \mathbf{A} \), \( \mathbf{A} \) being the non-random matrix

\[
A = \begin{bmatrix} f_1 & f_2 \\ s & 0 \end{bmatrix}
\]

with \( s := \frac{t_1 s^1 + p_1 s^2}{p_1 + t_1} \); \( f_i := \frac{t_i f^1_i + p_i f^2_i}{p_i + t_i} \), \( i = 1, 2 \).

The variables \( \eta_n \) are bounded away from zero and infinity and matrix \( \mathbf{A} \) is primitive, so \( A_{\text{erg}} \) is an ergodic set, and hypotheses A1–A3 of §4 are met.

Now we proceed to calculate the s.g.r. and the s.l.v. for the reduced system (5.4). In order to do so we will use the fact that the evolution of the population structure is a deterministic one and only the total number of individuals evolves, and we will reduce the problem to the scalar case, which we have already solved in a previous example. Specifically, let

\[
\mathbf{Y}_n = \bar{\mathbf{M}}_{\tau_n} \mathbf{Y}_{n-1} = \cdots = \eta_{\tau_n} \eta_{\tau_{n-1}} \cdots \eta_{\tau_1} \mathbf{A}^n \mathbf{Y}_0 = G_{\tau_n} \mathbf{A}^n \mathbf{Y}_0,
\]

where \( G_{\tau_n} := \eta_{\tau_n} \eta_{\tau_{n-1}} \cdots \eta_{\tau_1} \). Let \( \mu_1 \geq \mu_2 \) be the eigenvalues of \( \mathbf{A} \) and let \( \mathbf{u}_1, \mathbf{u}_2 \) be associated right eigenvectors. Since \( \mathbf{A} \) is primitive then \( \mu_1 > 0 \) and \( \mu_1 > \mu_2 \) and so \( \mathbf{A} \) can be diagonalized as \( \mathbf{A} = \mathbf{V} \mathbf{D} \mathbf{V}^{-1} \) with \( \mathbf{D} = \text{diag}(\mu_1, \mu_2) \), \( \mathbf{V} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix} \). Defining new stochastic variables \( \mathbf{S}_n = (\mathbf{S}^1_n, \mathbf{S}^2_n)^T = \mathbf{V}^{-1} \mathbf{Y}_n \), we have that \( \mathbf{S}_n = G_{\tau_n} \mathbf{D}^n \mathbf{S}_0 \) and, coming back to the \( \mathbf{Y}_n \),

\[
\mathbf{Y}_n = G_{\tau_n} (\mathbf{u}_1 \mu_1^n \mathbf{S}^1_0 + \mathbf{u}_2 \mu_2^n \mathbf{S}^2_0).
\]

As we mentioned before, the population structure \( \mathbf{Y}_n / \| \mathbf{Y}_n \| \) has a deterministic evolution and converges to \( \mathbf{u}_1 / \| \mathbf{u}_1 \| \). The logarithm of the total population \( \| \mathbf{Y}_n \| \) is

\[
\log \| \mathbf{Y}_n \| = \log G_{\tau_n} + \log \| \mathbf{u}_1 \mu_1^n \mathbf{S}^1_0 + \mathbf{u}_2 \mu_2^n \mathbf{S}^2_0 \|,
\]

where the second term in the right-hand side is deterministic. Now we apply to \( G_{\tau_n} \) the results that we obtained for the scalar system (5.1) so that \( (\log G_{\tau_n} - n\bar{\alpha}) / \sqrt{n} \sigma \xrightarrow{n \to \infty} \mathcal{N}(0, 1) \) where \( \bar{\alpha} := \mathbb{E}_\pi (\log \eta_{\tau_n}) \) and \( \sigma^2 \) is given by (5.3) with \( z_j := \log(\delta_j) - \bar{\alpha}, 1 \leq j \leq s \). Taking into account that

\[
\lim_{n \to \infty} \log \| \mathbf{u}_1 \mu_1^n \mathbf{S}^1_0 + \mathbf{u}_2 \mu_2^n \mathbf{S}^2_0 \| - \lim_{n \to \infty} \log(\mu_1^n \| \mathbf{u}_1 \mathbf{S}^1_0 \|) = 0
\]

we finally conclude that \( (\log \| \mathbf{Y}_n \| - an) / \sigma \sqrt{n} \xrightarrow{n \to \infty} \mathcal{N}(0, 1) \) where \( a = \bar{\alpha} + \log \mu_1 \).
(iv) Case 3

Let us assume that migration is deterministic, i.e. \( p_i(\zeta) = p_i \), \( t_i(\zeta) = t_i \) (with \( p_i, t_i \in (0, 1) \)) and that

\[
F_i^j(\tau_n) = f_i^j \eta_n; \quad S_i^j(\tau_n) = s_i^j; \quad i = 1, 2,
\]

where we assume that \( f_i^j, s_i^j \) are non-random, positive and \( s_i^j < 1 \). In this model survival is not affected by the environment, whereas fertility rates are. A good/bad environment increases/decreases fertility rates by the same multiplicative constant leaving the survival rates unaltered.

The matrices of vital rates in the reduced system are

\[
\bar{M}_{\tau_n} = \begin{bmatrix} f_1 \eta_n & f_2 \eta_n \\ s & 0 \end{bmatrix}
\]

where \( s = \frac{t_1 s_1^1 + p_1 s_2^2}{p_1 + t_1}; \quad f_i = \frac{t_i f_1^1 + p_i f_1^2}{p_i + t_i}, \quad i = 1, 2, \)

and we assume that \( \eta_n \) is a sequence of i.i.d. random variables such that \( 1/\eta_n \) has a probability density function \( g(x) \) given by

\[
g(x) = \begin{cases} 
0 & 0 < x < \varepsilon, \\
\gamma^\gamma \frac{\gamma - 1}{(\gamma - 1)!} x^{\gamma - 1} e^{-\gamma x} & \varepsilon \leq x \leq M, \\
0 & K < x,
\end{cases}
\]

where \( \varepsilon \) and \( K \) are arbitrary positive constants. Were it not for the fact that \( g(x) \) has been made 0 for \( x \) values smaller than \( \varepsilon \) and larger than \( K \), \( g(x) \) would be the probability density function of a gamma distribution used in Tuljapurkar (1990, p. 66) in the context of a two-age class model for which an exact distribution of age structure can be obtained. We use \( \varepsilon \) and \( C \) as cut-offs that make the support of \( g \) compact so that the elements of the \( M_{\zeta} \) (and with them those of \( \bar{M}_{\zeta} \)) are bounded away from zero and infinity, and so the set \( A_{ag} \) is ergodic. Since \( \varepsilon \) and \( K \) can be made, respectively, as small and large as we want, results in Tuljapurkar (1990) are, from a practical point of view, as good an approximation to our case as desired.

Thus, following Tuljapurkar (1990), we have that

\[
a = \log \frac{f_2 s}{f_1} - \mathbb{E}_C \left[ \log \left( \frac{f_2 y_n^2}{f_1 y_n^1} \right) \right],
\]

where \( C \) is the probability density function defined by

\[
C(x) = \kappa^{-1} x^{\gamma - 1} (1 + x)^{-\gamma} e^{-\gamma x}, \quad z = f_1^2 / (f_2 s),
\]

with \( \kappa \) being a normalizing constant.

This work was supported by Proyecto de Investigación MTM2008-06462 (MCYT).

Phil. Trans. R. Soc. A (2009)
Appendix

In the first place we will establish some results for the generic model (2.1) that will be useful further on. If we define

$$T_{n+1} = A_{r_{n+1}} A_{r_n} \cdots A_{r_1},$$

we can write (2.1) as $Z_n = T_n z_0$. Note that each realization of $T_n$ is a product of $n$ matrices belonging to the set $\mathcal{A}$ that we assume to be ergodic. The literature shows, Lemma A.1. (Equation (3′) in Tuljapurkar & Orzack (1980), lemma 3 in Furstenberg & Kesten (1960).) If $\mathcal{A}$ is an ergodic set with constants $g, \alpha$ and $\beta$ (see definition 2.1) then, for any realization of the process $T_n$, we have that, for $1 \leq i, j, l \leq N$,

$$\left| \frac{(T_{n+1})_{il}}{\sum_{q=1}^{N} (T_{n+1})_{iq}} - \frac{(T_{n+1})_{jl}}{\sum_{q=1}^{N} (T_{n+1})_{jq}} \right| \leq D_2 \eta^n, \quad (A1)$$

where $\eta < 1$ is a constant that can be written as

$$\eta = (1 - \hat{C}^{-3})^{1/g} \quad \text{with} \quad \hat{C}^2 = N^{2(g-1)} \left( \frac{\beta}{\alpha} \right)^{2g}, \quad (A2)$$

and $0 \leq D_2 < \infty$ is a constant that depends only on $N, g, \alpha$ and $\beta$.

Note that, if a set of matrices $\mathcal{A}_T$ is ergodic with constants $g, \alpha$ and $\beta$, so is the set of transposed matrices $\mathcal{A}_{T}^T$ so that equation (A1) is also valid if we interchange the subindexes of all the matrices $T_{n+1}$ that appear in the inequality. Therefore, from now on we will refer to equation (A1) to denote any one of these inequalities.

Using equation (A1) we will prove the so-called ‘weak ergodic theorem in demography’ (Cohen 1977, 1979; Tuljapurkar & Orzack 1980; Seneta 1981) for products of matrices belonging to an ergodic set, without resorting to Hilbert’s pseudometric. Moreover, we show that the speed with which vectors of population structure corresponding to two initial populations converge to each other depends only on the constants of the ergodic set. This fact will be necessary later on.

**Theorem A.2.** Let us assume that the set $\mathcal{A}$ is ergodic with constants $g, \alpha$ and $\beta$. Given $\delta > 0$ there exists $n_0$ such, that for all $n \geq n_0$, for all initial states $w_0, w_0' \in \mathcal{W}$ and for any realization of the process we have

$$\left\| W_n - W_n' \right\| = \left\| T_n w_0 - T_n w_0' \right\| \leq \delta, \quad (A3)$$

where, given $\delta$, the value of $n_0$ depends only on $N$ and on the constants of the ergodic set.

**Proof.** Given any $T_n, w_0$ and $w_0'$ we define $(T_n)_{*i} = \sum_{q=1}^{N} (T_n)_{qi} \neq 0$ for all $0 \leq i \leq N$ and we have

$$\frac{(T_n w_0)_k}{\left\| T_n w_0 \right\|} = \frac{\sum_{i=1}^{N} [(T_n)_{ki} / (T_n)_{*i}] (T_n)_{*i}(w_0)_i}{\sum_{j=1}^{N} (T_n)_{*j}(w_0)_j} = \sum_{i=1}^{N} (T_n')_{ki}(h_0)_i; \quad (A4)$$

Phil. Trans. R. Soc. A (2009)
with the definitions
\[
(T'_n)_{ki} := \frac{(T_n)_{ki}}{(T_n)_{ni}} \quad \text{so that} \quad \sum_{k=1}^{N} (T'_n)_{ki} = 1, \tag{A 5}
\]
\[
(h_0)_i := \frac{(T_n)_{si}(w_0)_i}{(\sum_{j=1}^{N} (T_n)_{sj}(w_0)_j)} \quad \text{and therefore} \quad \sum_{i=1}^{N} (h_0)_i = 1. \tag{A 6}
\]

Now we will prove that \(\|T'_n h_0 - T'_n h'_0\|\) reaches its maximum when \(h_0\) and \(h'_0\) are vectors belonging to the canonical basis of \(\mathbb{R}^N\) and therefore, according to (A 4)–(A 6), the first side of A 3 also reaches its maximum when \(w_0, w'_0\) belong to the canonical basis.

Denoting by \(t_1, \ldots, t_N\) the columns of \(T'_n\) and taking into account that \((h_0 - h'_0)_1 = -\sum_{j=2}^{N} (h_0 - h'_0)_j\) we have, whenever \((h_0 - h'_0)_1 \neq 0\), that
\[
\|T'_n h_0 - T'_n h'_0\| = \left| (h_0 - h'_0)_1 \right| t_1 \left( (h_0 - h'_0)_1 - (h_0 - h'_0)_1 \right) \sum_{j=2}^{N} t_j \left( (h_0 - h'_0)_j \right)_{1/1}.
\]

The norm in the right-hand side of this equation does not depend on the value of \((h_0 - h'_0)_1\), and only on the direction of the unitary vector with components \(-(h_0 - h'_0)_j/(h_0 - h'_0)_1\) with \(2 \leq j \leq N\), from where it follows that necessarily the maximum of \(\|T'_n h_0 - T'_n h'_0\|\) is reached either when \((h_0 - h'_0)_1 = 0\) or when \((h_0 - h'_0)_1 = \pm 1\). Reasoning along the same lines for the rest of the components \((h_0 - h'_0)_k\), we conclude that the maximum of \(\|T'_n h_0 - T'_n h'_0\|\) is reached when \((h_0 - h'_0)_k = 0, \pm 1, 1 \leq k \leq N\), i.e. \(h_0\) and \(h'_0\) must be two vectors of the canonical basis and the same happens to \(w_0\) and \(w'_0\).

Now let \(w_0 = e^j\) and \(w'_0 = e^k\) be two canonical vectors \((e^i)_i = \delta_{ij}, (e^k)_i = \delta_{ik}, 1 \leq i \leq N\). Then using equation (A 1) we have that for all \(w_0, w'_0 \in W'\)
\[
\left\| \frac{T_n w_0}{\|T_n w_0\|} - \frac{T_n w'_0}{\|T_n w'_0\|} \right\| \leq \left\| \frac{T_n e^j}{\|T_n e^j\|} - \frac{T_n e^k}{\|T_n e^k\|} \right\| = \sum_{q=1}^{N} \left| \frac{(T_n)_{qj}}{\sum_{p=1}^{N} (T_n)_{pj}} - \frac{(T_n)_{qk}}{\sum_{p=1}^{N} (T_n)_{pk}} \right| \leq N D_2 \eta^{n-1},
\]
where \(\eta\) is given by equation (A 2). Given \(\delta > 0\) the result follows taking \(n_0\) such that \(\delta = N D_2 \eta^{n_0-1}\).

Now let us consider the process \((D_n)_{n=1}^{\infty} := (\{\tau_n, W_n\})_{n=1}^{\infty}\) and its transition probabilities \(G_n\) defined by equation (2.3). We will denote \(G := G_1\). Under hypotheses (H1) and (H2), theorem 2.2 guarantees that the transition
probabilities $G_n \convergesto$ a probability distribution $F$ uniformly in the initial state of the chain. The proof of theorem 2.2 hinges on a series of lemmas, among which we have the following ones, which will be relevant thereafter.

**Lemma A.3.** (Lemma 5 in Cohen 1977.) Let us assume hypotheses (H1) and (H2). Given $\varepsilon > 0$ and $B \in \mathcal{B}$, $\exists \delta > 0$ such that for all $w, w' \in \mathcal{W}$ such that $\|w - w'\| \leq \delta$ and for all $\zeta \in \mathcal{I}$, $C \subset \mathcal{I}$ we have

$$|G(\zeta, w, C \times B) - G(\zeta, w', C \times B)| \leq \varepsilon.$$  

**Lemma A.4.** (Lemma 6 in Cohen 1977.) Let us assume hypotheses (H1) and (H2). Given $\varepsilon > 0$ and $B \in \mathcal{B}$, $\exists l_0 \in \mathbb{N}$ such that $\forall l \geq l_0$, $\forall \zeta \in \mathcal{I}$, $C \subset \mathcal{I}$ and $\forall w_0, w'_0 \in \mathcal{W}$ we have

$$|G(\zeta, w_0, C \times B) - G(\zeta, w'_0, C \times B)| \leq \varepsilon,$$

where $W_l, W'_l$ are obtained from $w_0, w'_0$ through equation (2.2).

Once we have established some results regarding general matrix models with environmental stochasticity, we will consider the original system (3.4) and the auxiliary system (3.7), so that

$$X_n^{(k)} = T_n^{(k)} x_0; \quad \hat{X}_n = \hat{T}_n x_0,$$

with

$$T_n^{(k)} := M_{\tau_0} P_{\tau_1} \cdots M_{\tau_{l-1}} P_{\tau_l}; \quad \hat{T}_n := M_{\tau_0} \hat{P}_{\tau_1} \cdots M_{\tau_{l-1}} \hat{P}_{\tau_l},$$

and we will introduce several results that relate these two systems.

As a direct consequence of proposition 4.1 we have, using theorem A.2,

**Proposition A.5.** Let us assume A2 and A3. Then, given $\delta > 0$ there exists $n_0 \in \mathbb{N}$ such that $\forall n \geq n_0, k \geq k_0$ ($k_0$ defined in proposition 4.1), $\omega \in \Omega$ and for any initial states $w_0, w'_0 \in \mathcal{W}$

$$\|W_n^{(k)} - W_n^{(k')\prime}\| = \left\| \frac{T_n^{(k)} w_0}{T_n^{(k)} w'_0} - \frac{T_n^{(k)} w'_0}{T_n^{(k)} w_0} \right\| \leq \delta.$$  

Moreover, given $\delta$ the value of $n_0$ depends only on $N$ and on the constants $g, \alpha$ and $\beta$ (defined in proposition 4.1).

Let us also consider the bivariate chains $\hat{W}_n = (\tau_n, \hat{W}_n)$ and $D_n^{(k)} = (\tau_n, W_n^{(k)})$, where $\hat{W}_n := \hat{X}_n/\|\hat{X}_n\|$ and $W_n^{(k)} := X_n^{(k)}/\|X_n^{(k)}\|$ and let their transition probabilities be defined through

$$\hat{G}_n(\zeta, w, C \times B) := \mathbb{P}\{\tau_n \in C, \hat{W}_n \in B \mid \tau_0 = i, \hat{W}_0 = w\},$$

$$G_n^{(k)}(\zeta, w, C \times B) := \mathbb{P}\{\tau_n \in C, W_n^{(k)} \in B \mid \tau_0 = i, W_0^{(k)} = w\},$$

with $\zeta \in \mathcal{I}$, $w \in \mathcal{W}$, $C \subset \mathcal{I}$ and $B \in \mathcal{B}$. We will denote $\hat{G} := \hat{G}_1$ and $G^{(k)} := G_1^{(k)}$.

If we assume hypotheses A1–A3, then, as a consequence of corollary 4.2, for all $k \geq k_0$ the chains $\hat{D}_n$ and $D_n^{(k)}$ meet the hypotheses of theorem 2.2, and so there exist stationary probability distributions $\hat{F}$ and $F^{(k)}$ for $\hat{D}_n$ and $D_n^{(k)}$, respectively. Specifically we have,
Lemma A.6. Given hypotheses A1–A3 we have,

(i) Given \( \varepsilon > 0, \ C \subset \mathcal{I} \) and \( B \in \mathcal{B}, \ \exists n_0 \in \mathbb{N}, \) such that \( \forall n \geq n_0, \ z \in \mathcal{I}, \ w \in \mathcal{W} \)

\[
|\hat{G}_n(z, w, C \times B) - \hat{F}(C \times B)| \leq \varepsilon.
\]

(ii) For all \( k \geq k_0 \) we have that, given \( \varepsilon > 0, \ C \subset \mathcal{I} \) and \( B \in \mathcal{B}, \) there exists \( n_0 \in \mathbb{N} \) (possibly depending on \( k \)) such that \( \forall n \geq n_0, \ z \in \mathcal{I}, \ w \in \mathcal{W} \)

\[
|G_n^{(k)}(z, w, C \times B) - F^{(k)}(C \times B)| \leq \varepsilon.
\]

Our goal now is to prove that \( F^{(k)} \) converges to \( \hat{F} \) when \( k \to \infty. \) In the first place we establish a number of relationships between the systems (3.4) and (3.7) when \( n \) is fixed and \( k \to \infty. \)

Lemma A.7. Let us assume hypothesis A1–A3. Then,

(i) Given \( n \in \mathbb{N}, \lim_{k \to \infty} \|T_n^{(k)} - \hat{T}_n\| = 0, \) uniformly in \( \omega \in \Omega. \)

(ii) Given \( n \in \mathbb{N}, \lim_{k \to \infty} \|W_n^{(k)} - \hat{W}_n\| = 0 \) uniformly in \( w_0 \in \mathcal{W} \) and \( \omega \in \Omega. \)

(iii) For all \( B \in \mathcal{B} \) and \( n \in \mathbb{N}, \lim_{k \to \infty} C_n^{(k)}(z, w, C \times B) = \hat{G}_n(z, w, C \times B) \)

uniformly in \( w \in \mathcal{W}, \ z \in \mathcal{I} \) and \( C \subset \mathcal{I}. \)

Proof. (i) It suffices to prove the result for \( n = 2. \) Let \( \|*\| \) be any matrix norm induced by a vector norm. Let us denote \( A^{(k)}_z := M_k \bar{P}_k \) and \( A_z := M_k \bar{P}_z. \) Then

\[
\|T_2^{(k)} - \hat{T}_2\| = \|A^{(k)}_{t_2} A^{(k)}_{t_1} - A_{t_2} A_{t_1}\| = \|A^{(k)}_{t_2} A^{(k)}_{t_1} - A^{(k)}_{t_2} A_{t_1} + A^{(k)}_{t_2} A_{t_1} - A_{t_2} A_{t_1}\|
\]

\[
\leq \|A^{(k)}_{t_2}\| \|A^{(k)}_{t_1} - A_{t_1}\| + \|A_{t_1}\| \|A^{(k)}_{t_2} - A_{t_2}\|
\]

\[
\leq \beta \|A_{t_1}\| + \|A_{t_2}\| \leq 2\beta \sup_{z \in \mathcal{I}} \|A^{(k)}_z - A_z\|
\]

\[
\leq 2\beta \sup_{z \in \mathcal{I}} \|M_z\| \sup_{z \in \mathcal{I}} \|P^{(k)}_z - \bar{P}_z\| \leq 2\beta K \sup_{z \in \mathcal{I}} \|P^{(k)}_z - \bar{P}_z\| \to 0,
\]

where we have used (3.6) and the bounds (4.1) and (4.2).

(ii) Let \( \|*\| \) denote, depending on the context, the 1-vector norm and its associated matrix norm in \( \mathbb{R}^{N \times N}. \) Since \( A_{aux} \) is an ergodic set, its matrices cannot have any zero columns, so that, using equation (4.2), \( \sum_{i=1}^{N} (M_z \bar{P}_z)_{ij} \geq \alpha \) and moreover \( \sum_{i=1}^{N} (\hat{T}_n)_{ij} \geq \alpha^n. \) Therefore,

\[
\|\hat{T}_n w_0\| = \sum_{j=1}^{N} \left( \sum_{i=1}^{N} (\hat{T}_n)_{ij} \right) (w_0)_j \geq \alpha^n.
\]

\[\text{(A 7)}\]
Now we can write
\[
\|W_n^{(k)} - \hat{W}_n\| = \left\| \frac{T_n^{(k)}w_0}{\|T_n^{(k)}w_0\|} - \frac{\hat{T}_n w_0}{\|\hat{T}_n w_0\|} \right\| \leq \frac{\|T_n^{(k)}w_0\|}{\|\hat{T}_n w_0\|} + \frac{\|T_n^{(k)}w_0 - \hat{T}_n w_0\|}{\|\hat{T}_n w_0\|} \leq 2\|T_n^{(k)} - \hat{T}_n\|/\alpha^n,
\]
where we have taken into account equation (A 7). The result now follows from (i).
(iii) Immediate taking into account (ii), the definition of $G_{n}^{(k)}(s, w, C \times B)$ and
\[\hat{G}_{n}(s, w, C \times B)\] and the fact that convergence with probability 1 implies
convergence in distribution. ■

The following two lemmas extend lemmas 5 and 6 in Cohen (1977) to the system (3.4), the results being uniform in $k$. Specifically,

Lemma A.8. Let us assume hypothesis A1–A3. Given $\varepsilon > 0$, $B \in \mathcal{B}$, $\exists k_1 > 0$

such that if $w, w' \in \mathcal{W}$ verify that $\|w - w'\| \leq \delta$ and $\forall k \geq k_1$, $\forall \xi \in \mathcal{I}$, $\forall C \subset \mathcal{I}$, then

\[|G^{(k)}(\xi, w, C \times B) - G^{(k)}(\xi, w', C \times B)| \leq \varepsilon.\]

Proof.
\[
|G^{(k)}(\xi, w, C \times B) - G^{(k)}(\xi, w', C \times B)|
\]
\[
\leq |G^{(k)}(\xi, w, C \times B) - \hat{G}(\xi, w, C \times B)| + |\hat{G}(\xi, w, C \times B) - \hat{G}(\xi, w', C \times B)|
\]
\[
+ |\hat{G}(\xi, w', C \times B) - G^{(k)}(\xi, w', C \times B)|
\]
and the result follows applying part (iii) of lemma A.7 with $n = 1$ to the first and
third terms in the right-hand side and lemma A.3 to the second term. ■

Lemma A.9. Let us assume hypotheses A1–A3. Given $\varepsilon > 0$, $B \in \mathcal{B}$, $\exists k_0, k_2 \in \mathbb{N}$, such that $\forall l \geq l_0$, $\forall k \geq k_2$, $\forall \xi \in \mathcal{I}$, $\forall C \subset \mathcal{I}$ and $\forall w_0, w'_0 \in \mathcal{W}$, we have
\[
|G^{(k)}(\xi, W_l^{(k)} C \times B) - G^{(k)}(\xi, W_l^{(k)'}, C \times B)| \leq \varepsilon. \hspace{1cm} (A 8)
\]

Proof. From lemma A.8 we have that $\forall k \geq k_1, \forall W_l^{(k)}, W_l^{(k)'} \in \mathcal{W}$ such that
\[
\|W_l^{(k)} - W_l^{(k)'}\| \leq \delta, \hspace{1cm} (A 8)
\]
holds. Moreover, from proposition A.5 we have that
\[
given \delta > 0, \exists l_0 \in \mathbb{N} such that \|W_l^{(k)} - W_l^{(k)'}\| \leq \delta, \forall w_0, w'_0 \in \mathcal{W}, \ k \geq k_0, \ l \geq l_0. \ The
result follows by taking $k_2 = \max(k_0, k_1)$. ■

From lemmas A.8 and A.9, and proceeding analogously to the proof of
theorems 2 and 3 of Cohen (1977) one can prove the following two results,
equivalent to theorems 2 and 3 of Cohen (1977) for the chains $(D_n^{(k)})_{n=1}$
uniformly in $k$:

Phil. Trans. R. Soc. A (2009)
Proposition A.10. Let us assume hypotheses A1–A3. Given \( \varepsilon > 0, C \subset \mathcal{I} \) and \( B \in \mathcal{B} \), \( \exists n_0 \in \mathbb{N} \) and \( k_2 \) given by lemma A.9 such that \( \forall k \geq k_2, \forall n \geq n_0, \zeta, \zeta' \in \mathcal{I}, \) \( w, w' \in \mathcal{W} \) one has

\[
|G_n^{(k)}(\zeta, w, C \times B) - G_n^{(k)}(\zeta', w', C \times B)| \leq \varepsilon.
\]

Proposition A.11. Let us assume hypotheses A1–A3. For all \( k \geq k_2 \) with \( k_2 \) given by lemma A.9 there exists a stationary probability distribution \( F^{(k)} \) such that given \( \varepsilon > 0, C \subset \mathcal{I} \) and \( B \in \mathcal{B} \), \( \exists n_0 \in \mathbb{N} \) such that \( \forall n \geq n_0, \zeta \in \mathcal{I}, w \in \mathcal{W} \)

\[
\sup_{k \geq k_2} |G_n^{(k)}(\zeta, w, C \times B) - F^{(k)}(C \times B)| \leq \varepsilon.
\]

As a consequence of the previous results, we finally have the result we were aiming at.

Theorem A.12. Let us assume hypotheses A1–A3. Given \( C \subset \mathcal{I} \) and \( B \in \mathcal{B} \) we have

\[
\lim_{k \to \infty} F^{(k)}(C \times B) = \hat{F}(C \times B).
\]

Proof. Taking any \( \zeta \in \mathcal{I}, w \in \mathcal{W}, C \subset \mathcal{I}, B \subset \mathcal{B} \) we can write

\[
|F^{(k)}(C \times B) - \hat{F}(C \times B)| \leq \sup_{k \geq k_2} |F^{(k)}(C \times B) - G_n^{(k)}(\zeta, w, C \times B)|
\]

\[
+ |\hat{G}_n(\zeta, w, C \times B) - \hat{F}(C \times B)|
\]

\[
+ |G_n^{(k)}(\zeta, w, C \times B) - \hat{G}_n(\zeta, w, C \times B)|.
\]

Now the result follows if we first take \( \lim_{k \to \infty} \) and use part (iii) of lemma A.7, and then take \( \lim_{n \to \infty} \) and use part (ii) of lemma A.6 and proposition A.11. 

Proof of theorem 4.3. Applying theorem 2.3, the hypotheses guarantee that we can define a s.g.r. and a s.l.v. for the auxiliary system (3.7) through \( \hat{a} := \lim_{n \to \infty} \mathbb{E} \log \|\hat{X}_n\|/n \) and \( \hat{\sigma}^2 := \lim_{n \to \infty} \mathbb{V}[\log \|\hat{X}_n\|]/n \). The definition of the global variables in (3.8) implies that \( \|\hat{X}_n\| = \|Y_n\| \) and therefore we have that \( a = \hat{a} \) and \( \sigma = \hat{\sigma} \). Consequently, all we have to do is to prove that \( \lim_{k \to \infty} a^{(k)} = \hat{a} \) and \( \lim_{k \to \infty} \sigma^{(k)} = \hat{\sigma} \).

Applying theorem 2.3 to the auxiliary and the original (for \( k \geq k_0 \)) systems, we know that

\[
\hat{a} = \mathbb{E}_{\hat{F}} \left[ \log \frac{(M_\tau \hat{P}_\tau \hat{W}_1)}{(\hat{W}_1)} \right]
\]

(A9)

and

\[
a^{(k)} = \mathbb{E}_{F^{(k)}} \left[ \log \frac{(M_\tau P_{\tau} W_1^{(k)})}{(W_1^{(k)})} \right]. \quad (A10)
\]

Phil. Trans. R. Soc. A (2009)
Moreover, we can define, for \( r = 1, 2, \ldots \)
\[
\hat{C}_r := \mathbb{E}_{\hat{F}} \left[ \left( \log \frac{(M_{T+1} \hat{P}_{T+1} \hat{W}_{T+1})_1}{(\hat{W}_{T+1})_1} - \hat{a} \right) \left( \log \frac{(M_{T+2} \hat{P}_{T+2} \hat{W}_{T+2})_1}{(\hat{W}_{T+2})_1} - \hat{a} \right) \right],
\]
\[
c_r^{(k)} := \mathbb{E}_{F^{(k)}} \left[ \left( \log \frac{(M_{T+1} P_{T+1}^k W_{T+1}^{(k)})_1}{(W_{T+1}^{(k)})_1} - a^{(k)} \right) \left( \log \frac{(M_{T+2} P_{T+2}^k W_{T+2}^{(k)})_1}{(W_{T+2}^{(k)})_1} - a^{(k)} \right) \right],
\]
where \( x_0 > 0 \) and we have
\[
\hat{\sigma}^2 = \hat{C}_1 + 2 \sum_{r=2}^{\infty} \hat{C}_r
\]
and
\[
(\sigma^{(k)})^2 = c_1^{(k)} + 2 \sum_{r=2}^{\infty} c_r^{(k)}.
\]
Denoting \( \hat{f}(\omega) := \log((M_{T+2} \hat{P}_{T+2} \hat{W}_{T+2})_1/(\hat{W}_{T+2})_1) = \log((\hat{T}_{T+2} x_0)_1/(\hat{T}_1 x_0)_1) \) the expression in brackets in (A 9) and \( f^{(k)}(\omega) := \log((M_{T+2} P_{T+2}^k W_{T+2}^{(k)})_1/(W_{T+2}^{(k)})_1) = \log((T_{T+2} x_0)_1/(T_{T+2}^{(k)} x_0)_1) \) the expression in brackets in (A 10) we have
\[
|a^{(k)} - \hat{a}| = \left| \int_{\Omega} f^{(k)} dF^{(k)} - \int_{\Omega} \hat{f} d\hat{F} \right| \leq \left| \int_{\Omega} f^{(k)} dF^{(k)} - \int_{\Omega} \hat{f} d\hat{F} \right| + \left| \int_{\Omega} \hat{f} d\hat{F} - \int_{\Omega} \hat{f} d\hat{F} \right| \leq \sup_{\omega \in \Omega} \left| f^{(k)} - \hat{f} \right| + \left| \int_{\Omega} \hat{f} d\hat{F} - \int_{\Omega} \hat{f} d\hat{F} \right|.
\]
Now we will show that both terms in the right-hand side of (A 13) tend to zero when \( k \to \infty \).

In the first place we will prove that \( \lim_{k \to \infty} \sup_{\omega \in \Omega} |f^{(k)} - \hat{f}| = 0 \). Let us denote \( H^{(k)}(\omega) := (T_{T+2}^{(k)} x_0)_1, \hat{H}(\omega) := (\hat{T}_{T+2} x_0)_1, L^{(k)}(\omega) := (T_1^{(k)} x_0)_1, \hat{L}(\omega) := (\hat{T}_1 x_0)_1 \) so that \( f^{(k)}(\omega) = \log(H^{(k)}(\omega)/L^{(k)}(\omega)) \) and \( \hat{f}(\omega) = \log(\hat{H}(\omega)/\hat{L}(\omega)) \). Using the bounds (4.2) we have that, for all \( \omega \in \Omega \),
\[
\hat{L}(\omega) \geq \alpha \min x_0; \quad \hat{\omega}(\omega) \leq \beta \max x_0; \quad \hat{H}(\omega) \geq \alpha^2 \min x_0; \quad \hat{H}(\omega) \leq \beta^2 \max x_0.
\]
Moreover, from part (i) of lemma A.7 it follows that \( \lim_{k \to \infty} H^{(k)}(\omega) = \hat{H} \) uniformly in \( \omega \in \Omega \) and \( \lim_{k \to \infty} L^{(k)} = \hat{L} \) uniformly in \( \omega \in \Omega \). Now we make use of the result by which if two sequences of functions \( M^{(k)}(\omega) \) and \( S^{(k)}(\omega) \) converge when \( k \to \infty \) to \( M(\omega) \) and \( S(\omega) \) uniformly in \( \omega \in \Omega \) and, moreover, for all \( \omega \) \( |M(\omega)| \leq C \) and \( |S(\omega)| \geq \delta \) for certain positive constants \( C \) and \( \delta \), then \( \lim_{k \to \infty} M^{(k)}(\omega)/S^{(k)}(\omega) = M(\omega)/S(\omega) \) uniformly on \( \omega \in \Omega \) (*) . As a consequence we have that \( \lim_{k \to \infty} H^{(k)}(\omega)/L^{(k)}(\omega) = \hat{H}(\omega)/\hat{L}(\omega) \) uniformly on \( \omega \in \Omega \). Now, using (A 14) we have that there exist positive constants \( 0 < \varepsilon < K \).
such that $\varepsilon \leq \hat{H}(\omega)/\hat{L}(\omega) \leq K$ for all $\omega \in \Omega$. Since the logarithm is a uniform continuous function in $[\varepsilon, K]$ it follows, using (*), that $\lim_{k \to \infty} \sup_{\omega \in \Omega} |f^{(k)} - \hat{f}| = 0$ as we wanted to show.

From the discussion above it follows that $\hat{f}(\omega) \leq L$ for a certain constant $L > 0$. Now using this fact, theorems 5 and 25.8 in Billingsley (1986), it follows that the second term in the right-hand side of equation (A13) tends to zero when $k \to \infty$ and so (4.3) holds.

Now let $r \in \mathbb{N}$ be fixed. Proceeding in the same way as above it follows that

$$\lim_{k \to \infty} c_r^{(k)} = \hat{C}_r. \tag{A15}$$

We will show that the series in the right-hand side of equation (A11) converges uniformly in $k$ for $k \geq k_0$. Expressions 3.26 and 3.27 in Furstenberg & Kesten (1960) applied to the auxiliary system show that $\hat{C}_r = O(\lambda_r^2)$ with $\lambda_2 = \{\max(1 - C^{-3}, \lambda_1) \}^{1/2}$, where $C > 1$ is defined in equation (3.2) of that reference and $\lambda_1 < 1$ in the hypothesis $AII$ (p. 464). Therefore, the series that defines $\hat{\sigma}_r^2$ converges.

Reasoning in the same way for the original system we have that for all $k \geq k_0$

$$c_r^{(k)} = O(\lambda_r^2) \quad \text{where} \quad \lambda_2 = \{\max(\eta, \rho) \}^{1/2},$$

where $\eta$ is given in lemma A.1 when applied to system (3.4) (recall that, owing to proposition 4.1, the constants of the ergodic set $A_r^{(k)}$ can be chosen independent of $k$ and therefore $\eta$ is independent of $k$) and $\rho$ in (2.4). Therefore, the series in the right-hand side of (A12) converges uniformly in $k$ for $k \geq k_0$. Therefore taking the limit when $k \to \infty$ in (A12), interchanging the series with the limit due to the uniform convergence of the series, and taking into account (A15) we obtain (4.4) as desired.

References


Phil. Trans. R. Soc. A (2009)


