Approximate reduction of linear population models governed by stochastic differential equations: application to multiregional models

Luis Sanz and Juan Antonio Alonso

Dpto. Matemática del Área Industrial, ETSI Industriales, Univ. Politécnica de Madrid, Madrid, Spain

ABSTRACT
In this work we develop approximate aggregation techniques in the context of slow-fast linear population models governed by stochastic differential equations and apply the results to the treatment of populations with spatial heterogeneity. Approximate aggregation techniques allow one to transform a complex system involving many coupled variables and in which there are processes with different time scales, by a simpler reduced model with a fewer number of ‘global’ variables, in such a way that the dynamics of the former can be approximated by that of the latter. In our model we contemplate a linear fast deterministic process together with a linear slow process in which the parameters are affected by additive noise, and give conditions for the solutions corresponding to positive initial conditions to remain positive for all times. By letting the fast process reach equilibrium we build a reduced system with a lesser number of variables, and provide results relating the asymptotic behaviour of the first- and second-order moments of the population vector for the original and the reduced system. The general technique is illustrated by analysing a multiregional stochastic system in which dispersal is deterministic and the rate growth of the populations in each patch is affected by additive noise.

1. Introduction
Nature offers many examples of systems with an inherent complexity whose study leads to mathematical models with a large number of state variables whose analytical study is, in most cases, not feasible. In order to be able to extract important information about the behaviour of some of these complex models, one can resort to ‘approximate aggregation methods’. These are mathematical techniques, which are usually applied in systems governed by processes with different time scales, in which appropriate approximations are introduced in order to transform the system under consideration into a reduced system with a lower number of variables, called ‘global variables’. In this way, the behaviour of the original system can be approximated, but not known with exactitude, in terms of the knowledge of the behaviour of the reduced system.

CONTACT Luis Sanz, lsanz@upm.es; Juan Antonio Alonso, jalonso@etsii.upm.es

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Approximate aggregation techniques in population dynamics have been widely studied in the context of deterministic systems with different time scales both in continuous and discrete time (see the review in [2] and the references therein), as well as in discrete stochastic models incorporating either environmental [19] or demographic stochasticity [21].

Stochastic differential equations (SDEs) can be thought of as resulting from the introduction of environmental stochasticity in the coefficients of deterministic ODEs. In spite of the difficulty of their analytical treatment, they have become popular as population modeling tools (see for example [6, 8] for the scalar case and [13, 24] for applications to competition models and epidemiology respectively).

The aim of this work is to formulate an approximate aggregation technique valid to reduce fast-slow linear population models governed by SDEs, to give sufficient conditions for the solutions of the models to be positive and to relate the asymptotic behaviour of the first- and second-order moments of the population vector for the original and the reduced system. The original model is built by considering a fast deterministic process that converges to an equilibrium, together with a slow process in which the parameters are affected by additive noise. In the resulting system the quotient between the rate of diffusion squared and the speed of drift is \( O(\varepsilon) \), \( \varepsilon \to 0 \), where \( \varepsilon \) is a measure of the difference of time scales between the fast and the slow process. This is in contrast with most previous works in the field [5, 10, 11, 16] which are valid in other situations. An exception is [14] which covers our case of interest but which only deals with the relationships between the original and the reduced systems in finite time intervals.

The existing literature on the field of the approximate reduction of SDEs has been developed mainly in the context of physics and control theory. Our approach, although general in nature, is aimed to population dynamics applications. Specifically, we will employ it to study a multiregional model consisting of a single population living in a multipatch environment in such a way that dispersal is deterministic and the growth rate of the population in each patch is affected by additive noise. Models with spatial heterogeneity have been widely studied through aggregation techniques (see [3] for the continuous time case and [2] for discrete time), normally making use of the fact that in many practical situations the dispersal of individuals amongst the different patches is fast with respect to other processes like demography or competition (an exception to this is [20]). In this work we contemplate a situation that has not been dealt with so far: dispersal between some spatial patches can be fast with respect to demography whereas migration between some other patches can happen at the same time scale of demography. For example we can think of a population of birds living in different islands in such a way that inter-island movements happen at the scale of demography but, in comparison, intra-island migrations are fast.

The manuscript is organized as follows: in Section 2 we present the general formulation of linear SDE models and, in order to be able to use them in population dynamics applications, we give sufficient conditions that guarantee that if initial conditions are positive, the solution to such a system remains positive for all times. Section 3 presents the general formulation of a linear two time scale population model with stochasticity affecting the slow process. Section 4 carries out the reduction of the system, which is accomplished by letting the fast process reach equilibrium and defining and adequate set of new variables. In Section 5 we apply the general technique to the case of a stochastic multiregional model, with special attention to the case in which all migrations are fast with respect to demography, since it is the most frequent situation in practice and moreover in that case...
the reduced system is scalar. Section 6 presents a result that allows one to relate the asymptotic behaviour of the first and second statistical moments of the population vector for the original and the reduced system, and applies it to study the asymptotic behaviour of the multiregional models of Section 5. A brief discussion of results and the Appendix with mathematical proofs complete the manuscript.

2. Linear population models with two time scales

Throughout the paper we assume we are working in a complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ where the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ satisfies the usual conditions [18]. Let us consider a structured population modelled by an linear autonomous homogeneous stochastic differential equation. The model has the form

$$dX(t) = AX(t) \, dt + \sum_{h=1}^{m} B^h X(t) \, dW^h(t), \quad X(0) = X_0, \quad (1)$$

where $X(t) = (X^1(t), X^2(t), \ldots, X^d(t)) \in \mathbb{R}^d$ is the population vector, $A = [a_{ij}] \in \mathbb{R}^{d \times d}$, $B^h = [b^h_{ij}] \in \mathbb{R}^{d \times d}$, $h = 1, 2, \ldots, m$ and $W(t) = (W^1(t), W^2(t), \ldots, W^m(t))^T$ is a $m$-dimensional standard Wiener process defined in the previous probability space. Moreover, we assume that $X_0 \in \mathbb{R}^d$ is a non-random vector. Models of the kind (1) are obtained from a linear deterministic model

$$\frac{dX(t)}{dt} = AX(t), \quad X(0) = X_0 \quad (2)$$

if we add noise to the population vital rates $a_{ij}$. Indeed, system (1) can be written in the form

$$dX^i(t) = \sum_{j=1}^{d} (a_{ij} \, dt + b^1_{ij} \, dW^1(t) + \cdots + b^m_{ij} \, dW^m(t)) X^j(t), \quad i = 1, \ldots, d$$

from where we see that each coefficient $b^h_{ij}$ characterizes the intensity of the noise $dW^h$ affecting $a_{ij}$. Note that the case in which the noises $dW^h(t)$ are correlated can be reduced to this setting through an appropriate transformation [1, p. 126]. Systems of the kind (1) can be interpreted in the sense of Ito or in the sense of Stratonovich, and the results obtained in the two cases differ [23]. However, one can choose any of the two interpretations as long as one defines appropriately the parameters of the model [7]. In this work we will make use of the Ito interpretation. It is well known [1, p. 126] that in the previous conditions there exists a unique solution to (1) which is continuous with probability one.

In order to be a valid model for a population, the solution of (1) for any positive initial condition must remain non-negative for all times, so we turn our attention to this property. Given a vector or matrix $x$ we will write $x \geq 0$ (resp. $x > 0$) to denote that all the components of $x$ are non-negative (resp. positive). Let us recall that a square matrix $A$ is said to be a Metzler matrix or a essentially non-negative matrix when it has non-negative off-diagonal entries, that is, $a_{ij} \geq 0$ for $i \neq j$ [22]. It is well known that if $A$ is a Metzler matrix then solutions of the deterministic system (2) meet the desired property. However,
in model (1) additional requirements are needed in order to ensure positivity of solutions. The next result gives sufficient conditions:

**Theorem 2.1:** If matrix $A$ is a Metzler matrix and matrices $B^1, \ldots, B^k$ are diagonal, then given $X(0) = X_0 > 0$, the solution of (1) verifies that $P(X(t) > 0 \text{ for all } t \geq 0) = 1$.

**Proof:** See Appendix. ■

3. A model with two time scales

We suppose a stage-structured population in which individuals are classified into stages or groups attending to any characteristic of the life cycle. Moreover, each of these groups is divided into several subgroups that can correspond to different spatial patches, different individual activities or any other characteristic that could change the life cycle parameters. The model is therefore general in the sense that we do not state in detail the nature of the population or the subpopulations.

We consider the population being subdivided in $q$ populations (or groups). Each group is subdivided in subpopulations (subgroups) in such a way that for each $i = 1, 2, \ldots, q$, group $i$ has $N_i$ subgroups. Therefore, the total number of subgroups is $N := N_1 + N_2 + \cdots + N_q$. We denote the fast time as $\tau$, while the slow time will be denoted by $t$. In this way we have $t = \varepsilon \tau$ where $\varepsilon << 1$ is a small number that represents the ratio between the slow and the fast times.

Let $x_{ij}(\tau)$ be the density of subpopulation $j$ of population $i$ at time $\tau$, with $i = 1, 2, \ldots, q$ and $j = 1, 2, \ldots, N_i$. In order to describe the population of group $i$ at time $\tau$ we will use vector $x^i(\tau) = (x^1_i(\tau), x^2_i(\tau), \ldots, x^{N_i}_i(\tau))^T \in \mathbb{R}^{N_i}$, $i = 1, 2, \ldots, q$, where $T$ denotes transposition. The composition of the total population is then given by vector $X(\tau) = (x^1(\tau)^T, x^2(\tau)^T, \ldots, x^1(\tau)^T)^T \in \mathbb{R}^N$.

In the evolution of the population we will consider two linear processes whose corresponding characteristic time scales, are very different from each other. In order to include in our model both time scales we will model these two processes, to which we will refer as the fast and the slow dynamics, by two different matrices.

In principle, we will make no special assumptions regarding the characteristics of the slow dynamics other than linearity and restrictions to guarantee positivity of solutions to the slow process. Thus, we will assume that the parameters of the slow process are defined by

$$S + \sum_{h=1}^{k} B^h \frac{dW^h(t)}{dt},$$

where: $S.1.$ Matrix $S \in \mathbb{R}^{N \times N}$ models the deterministic part of the slow process and has non-negative off-diagonal entries, that is, $S$ is a Metzler matrix. We consider $S$ divided into blocks $S_{ij}$, $1 \leq i, j \leq q$ in such a way that

$$S = \begin{bmatrix} S_{11} & \cdots & S_{1q} \\ \vdots & \ddots & \vdots \\ S_{q1} & \cdots & S_{qq} \end{bmatrix}.$$
Each block $S_{ij} = [s_{ij}^{\alpha\beta}]$ has dimensions $N_i \times N_j$ and characterizes the rates of transference of individuals from the subgroups of group $j$ to the subgroups of group $i$. More specifically, for each $\alpha = 1, 2, \ldots, N_i$ and each $\beta = 1, 2, \ldots, N_j$, entry $s_{ij}^{\alpha\beta}$ represents the rate of transference of individuals from subgroup $\beta$ of group $j$ to subgroup $\alpha$ of group $i$.

S.2. $W^1, \ldots, W^k$ are independent standard Wiener processes (so that by $dW^1(t)/dt, \ldots, dW^k(t)/dt$ we denote the associated weakly defined gaussian white noises) and, for each $h = 1, \ldots, k$, matrix $B^h \in \mathbb{R}^{N \times N}$ models the contribution of noise $dW^h(t)$ to the dynamics of the slow process. In order to be able to guarantee positivity of solutions (Theorem 2.1), we assume that matrices $B^h$ are diagonal, that is,

$$B^h = \text{diag}(B^h_1, \ldots, B^h_q), \quad h = 1, \ldots, k \quad \text{with} \quad B^h_i = \text{diag}(b^h_{i1}, \ldots, b^h_{iN_i}), \quad i = 1, \ldots, q.$$  

Note from Equation (3) that $b^h_{i\alpha}$ models the intensity of the noise $dW^h(t)$ affecting coefficient $s^{\alpha\alpha}_{ii}$.

In this context, we say that an eigenvalue $\lambda$ of a certain square matrix $A$ is strictly dominant when the real part of $\lambda$ is strictly larger than the real part of the rest of the eigenvalues of $A$.

As far as the behaviour of the fast dynamics is concerned, we will make the following three assumptions:

F.1. The fast process is deterministic.

F.2. The fast dynamics is an internal process for each group, that is, there is no transference of individuals from one group to a different one. Therefore, for each $i = 1, \ldots, q$, the fast dynamics of group $i$ will be represented by a Metzler matrix $F_i$ of dimensions $N_i \times N_i$. We will assume that $F_i$ is irreducible in the sense that there exists $r > 0$ such that $rI + F_i$ is a primitive non-negative matrix [22]. Therefore, the matrix that governs the fast dynamics for the whole population is

$$F = \text{diag}(F_1, F_2, \ldots, F_q).$$  

F.3. The fast process in each group has a non-trivial equilibrium point which is asymptotically stable. More specifically, for each $i = 1, \ldots, q$ matrix $F_i$ has eigenvalue $0$ and it is strictly dominant so that the rest of the eigenvalues of $F_i$ have negative real parts. Since $F_i$ is a Metzler irreducible matrix, eigenvalue $0$ is simple [22, Theorem 2.6] and moreover, there exist positive right and left eigenvectors $v_i$ and $u_i$ of $F_i$ associated to eigenvalue $0$ for which we choose the following normalization conditions

$$F_i v_i = 0; \quad u_i^T F_i = 0^T$$  

$$\|v_i\|_1 = 1; \quad u_i^T v_i = 1,$$  

where $\|*\|_1$ denotes the 1-vector norm.

In order to incorporate both time scales in our model, we will make use of parameter $\varepsilon = t/\tau$. The model, to which we will refer as original system, has the following form in the slow time

$$dX(t) = \frac{1}{\varepsilon}FX(t) \, dt + SX(t) \, dt + \sum_{h=1}^k B^h(t) \, dW^h(t).$$
Alternatively, using the fast time $\tau$ we have

$$dX(\tau) = FX(\tau) \, d\tau + \varepsilon SX(\tau) \, d\tau + \sqrt{\varepsilon} \sum_{h=1}^{k} B^h X(\tau) \, dW^h(\tau),$$

(7)

where we have made an abuse of notation and kept the same notation for $X$ and the $W^h$ in the new time, and we have used [1, p. 47] that if $W(t)$ is a standard Wiener process then so is $\sqrt{\varepsilon} W(t/\varepsilon)$. We stress that the quotient between the rate of diffusion squared and the speed of drift is $O(\varepsilon)$, $\varepsilon \rightarrow 0$, which is in contrast with most approaches in the analysis of two-time scales systems [5, 10, 11, 16] which are valid in other situations.

4. Approximate reduction of the model

In order to reduce the original system (7), we will use the fact that the fast process has an asymptotic stable equilibrium, and we will approximate this system by another one in which the fast process has reached equilibrium.

Let $i = 1, \ldots, q$ be fixed and let $\hat{F}_i := v_i u_i^T$. From Equation (5), we have that if the system were governed by the fast process exclusively, for any initial condition $x^i \in \mathbb{R}_{+}^N$, the population of each group $i$ would tend to vector $\hat{x}^i := \hat{F}_i x^i = (u_i^T x^i) v_i$. From this expression we note that vector $v_i$ defines the equilibrium population structure for group $i$ and $u_i$ is a vector of reproductive values, that is, the larger $u_i^j$, the higher the contribution of the $j$-th subgroup of the $i$th group to the equilibrium population. Therefore $u_i$ characterizes the size of the equilibrium population.

We define matrix $\hat{F}$ in the following way,

$$\hat{F} = \text{diag}(\hat{F}_1, \hat{F}_2, \ldots, \hat{F}_q),$$

and then for any initial condition $X \in \mathbb{R}_{+}^N$, the equilibrium population for the fast dynamics in the whole system is given by $\hat{X} = \hat{F}X$. Let us define the non-negative matrices

$$V := \text{diag}(v_1, v_2, \ldots, v_q) \in \mathbb{R}^{N \times q}, \quad U := \text{diag}(u_1^T, u_2^T, \ldots, u_q^T) \in \mathbb{R}^{q \times N}$$

whose interpretation is immediate bearing in mind what we pointed out about $v_i$ and $u_i$.

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

**Lemma 4.1:** Matrices $F, \hat{F}, V$ and $U$ verify:

(a) $\hat{F} = UV$, $UV = I_q$ and the columns of $V$ are independent and constitute a basis of $\text{Im} \hat{F}$.
(b) $\hat{F} \hat{F} = 0$.

Now, from Equation (6) we will build an auxiliary system replacing the state variables in the right side by its equilibrium values for the fast process and use that $\hat{F} \hat{F} = 0$,

$$dX(t) = \frac{1}{\varepsilon} \hat{F} FX(t) \, dt + SFX(t) \, dt + \sum_{h=1}^{k} B^h \hat{F}X(t) \, dW^h(t)$$
\[ \dot{X}(t) = \bar{S}Y(t) \, dt + \sum_{h=1}^{k} \bar{B}^h Y(t) \, dW^h(t). \] (8)

Now we define the vector of global variables as

\[ Y = (y^1, \ldots, y^q)^T : = UX \in \mathbb{R}^q \] (9)

and, multiplying Equation (8) on the left by \( U \) and using that \( \hat{F} = VU \) and Equation (9) we obtain the aggregated system

\[ \dot{Y}(t) = \bar{S}Y(t) \, dt + \sum_{h=1}^{k} \bar{B}^h Y(t) \, dW^h(t), \] (10)

where we have defined

\[ \bar{S} = [\bar{s}_{ij}] := USV = \begin{bmatrix} u_1^T S_{11} v_1 & \cdots & u_q^T S_{1q} v_q \\ \vdots & \ddots & \vdots \\ u_1^T S_{q1} v_1 & \cdots & u_q^T S_{qq} v_q \end{bmatrix} \in \mathbb{R}^{q \times q} \] (11)

\[ \bar{B}^h : = UB^h V = \text{diag} \left( \sum_{\alpha=1}^{N_1} u_1^e \alpha u_1^b \bar{b}^h_{1\alpha}, \ldots, \sum_{\alpha=1}^{N_q} u_q^e \alpha u_q^b \bar{b}^h_{q\alpha} \right) \in \mathbb{R}^q, \quad h = 1, \ldots, k. \] (12)

Note that the global variables \( Y(t) \) defined by Equation (9) have the following expression in terms of the variables \( X(t) \) of the auxiliary system:

\[ y^i(t) = u_i^T x^i(t) = u_1^i x^{i1}(t) + u_2^i x^{i2}(t) + \cdots + u_{Ni}^i x^{Ni}(t), \quad i = 1, \ldots, q. \]

Note that:

(a) \( y^i(t) \) is a linear combination of the variables corresponding to group \( i \), being the coefficients of the combination the components of vector \( u_i \). Recall that \( u_i \) is a vector of reproductive values for the fast process in group \( i \). Therefore, for each \( j = 1, \ldots, N_i \), variable \( x^{ij}(t) \) has a relative weight in \( y^i(t) \) which is proportional to \( u_j^i \), that is, proportional to the contribution to the total equilibrium population that an individual initially present in group \( i \) and subgroup \( j \) would have in the case that the system were governed by the fast process exclusively.

(b) The global variables are conservative for the fast process. Indeed, suppose that the fast process is the only one acting in the system. Then we would have \( \dot{X}(t) = FX(t)/\varepsilon \) and using that \( UF = 0, \dot{Y}(t) = UX(t) = UFX(t)/\varepsilon = 0 \).

(c) The components of the matrices representing the drift and the diffusion for the reduced system are certain linear combinations of their analogues for the original system, where the coefficients of the combination are determined by the equilibrium characteristics of the fast process.

The next result together with Theorem 2.1 guarantees that the original and the aggregated systems have positive solutions for any positive initial conditions.
Proposition 4.2: \( F + \varepsilon S \) and \( \bar{S} \) are Metzler matrices and matrices \( \bar{B}^h, h = 1, \ldots, k \) are diagonal. Therefore, according to Theorem 2.1 both the original system (7) and the aggregated system (10) verify that for any positive initial condition the solution remains positive for all \( t > 0 \) with probability one.

**Proof:** \( F + \varepsilon S \) is clearly a Metzler matrix for it is the sum of Metzler matrices. Now let \( i \neq j \). Since \( S \) is a Metzler matrix we have that \( S_{ij} \geq 0 \) and using the fact that \( u_i \) and \( v_j \) are positive vectors, from Equation (11) it follows that \( \bar{s}_{ij} = u_i^T S_{ij} v_j \geq 0 \) and so \( \bar{S} \) is a Metzler matrix. Moreover, from Equation (12) it is clear that \( \bar{B}^h \) is a diagonal matrix for each \( h = 1, \ldots, k \). \( \blacksquare \)

5. Multiregional models with two time scales

In this section we will illustrate the reduction technique by applying it to a multiregional model.

5.1. Model setting

We consider a population living in a multipatch system. We assume that there are a number \( N \) of different patches among which the individuals can migrate. The growth of the population in each patch is linear and is affected by stochasticity. Migration among patches is assumed to be deterministic.

We number the patches in the form \((i, \alpha)\), \(i = 1, 2, \alpha = 1, \ldots, N_i\), where \( N := N_1 + N_2 \). Coming back to the terminology of Section 3, the first index, \( i \), defines the ‘group’ of patches and the second, \( \alpha \), the ‘subgroup’ within that group. Note that in our setting we have chosen that the number \( q \) of groups is 2 just for the sake of simplicity in the expression of the matrices involved, but the generalization of the model to an arbitrary number \( q \) of groups is straightforward. Let \( x^{ij}(t) \) denote the population in group \( i \) and subgroup \( j \) at time \( t \) and let \( X = (x^{11}, \ldots, x^{1N_1}, x^{21}, \ldots, x^{2N_2})^T \in \mathbb{R}^N \) be the population vector.

We assume that migration is fast within each group of patches, that is, from any patch \((i, \alpha)\) to any other patch of the form \((i, \beta)\), \( \beta \neq \alpha \). Migration is assumed to be slow between patches belonging to different groups, that is, from any patch \((i, \alpha)\) to any other patch of the form \((j, \beta)\), with \( j \neq i \). We can think of each group of patches as spatial regions located close to each other so that migration between them is easy, whereas different groups of patches correspond to regions amongst which migration is more difficult. For example, in a population of birds each group of patches can correspond to an island and the subgroups can correspond to the different spatial locations within an island, so that intra-island migrations are fast with respect to inter-island movements.

Let \( \tau \) and \( t \) denote, respectively, the times corresponding to the fast and the slow migrations and let \( \varepsilon := \tau / t \) be the ratio between both. We assume that the growth of the population takes place in the slow time scale. For each pair \((i, \alpha)\), \( i = 1, 2, \alpha = 1, \ldots, N_i \), let \( r_i^\alpha \) be the deterministic population growth rate in patch \((i, \alpha)\) and let us assume that this growth rate is affected by a noise defined by a certain linear combination \( \sigma_{1i\alpha}^1 dW^1(t)/dt + \cdots + \sigma_{ki\alpha}^k dW^k(t)/dt \) of (weakly defined) independent white noise
Regarding the slow migration between different groups of patches, for each \( i \neq j \) and each \( \alpha = 1, \ldots, N_i, \beta = 1, \ldots, N_j \), we define \( t_{ij}^{\alpha \beta} \geq 0 \) as the (slow) migration coefficient from patch \((j, \beta)\) to patch \((i, \alpha)\). Similarly we define \( t_{ij}^{\alpha \beta} = 0 \) for each \( \alpha, \beta = 1, \ldots, N_i \) with \( \alpha \neq \beta \) (as there is no slow migration within a group of patches) and

\[
t_{ij}^{\alpha \alpha} := - \sum_{j=1, j \neq i=1, \beta \neq \alpha}^{q} \sum_{j=1, j \neq i=1, \beta \neq \alpha}^{N_i} t_{ij}^{\beta \alpha}, \quad i = 1, 2, \quad \alpha = 1, \ldots, N_i.
\]

Let us define matrices \( L_{ij} := [t_{ij}^{\alpha \beta}] \in \mathbb{R}^{N_i \times N_i}, i, j = 1, \ldots, q \) and

\[
L := \begin{bmatrix}
L_{11} & L_{12} \\
L_{21} & L_{22}
\end{bmatrix}.
\]

Then the slow process, that is, the joint effect of the growth process and of the slow migration between patches of different groups, can be modelled by the following system of SDEs

\[
dX(t) = (L + R)X(t) \, dt + \sum_{h=1}^{k} G^h X(t) \, dW^h(t), \quad X(0) = X_0.
\]

Regarding the fast migration between patches of the same group, for each \( i = 1, 2 \) and \( \alpha, \beta = 1, \ldots, N_i \) with \( \alpha \neq \beta \), let \( m_{i}^{\alpha \beta} \geq 0 \) be the (fast) migration rate from patch \((i, \beta)\) to patch \((i, \alpha)\). For \( \beta = \alpha \) we define

\[
m_{i}^{\alpha \alpha} := - \sum_{\beta=1, \beta \neq \alpha}^{N_i} m_{i}^{\beta \alpha}.
\]

(13)

Now let \( M_i := [m_{i}^{\alpha \beta}] \in \mathbb{R}^{N_i \times N_i}, i = 1, 2 \). We assume that the \( m_{i}^{\alpha \beta} \) are such that \( M_i \) is irreducible for each \( i = 1, 2 \). From Equation (13) we have that the columns of each \( M_i \) add up to zero and so matrix \( M_i + I \) is a non-negative primitive column stochastic matrix. Therefore 0 is the (strictly) dominant eigenvalue of \( M_i \) and moreover it is simple. Let \( u_i := 1_i = (1, \ldots, 1)^T \in \mathbb{R}^{N_i} \) and \( v_i = (v_1^i, \ldots, v_{N_i}^i)^T \in \mathbb{R}^{N_i} \) be its associated positive left and right eigenvectors, where we assume the normalization condition \( 1_i^T v_i = 1 \). Note that vector \( v_i \) defines the equilibrium distribution between the different patches of group \( i \) when we consider the fast migration as the only process acting on the system. Now we define matrices

\[
M := \text{diag}(M_1, M_2) \in \mathbb{R}^{N \times N}
\]
\[ V := \text{diag}(v_1, v_2) \in \mathbb{R}^{N \times 2}, \quad U := \text{diag}(1_1^T, 1_2^T) \in \mathbb{R}^{2 \times N} \]
\[ \hat{M}_i := v_i 1_i^T \in \mathbb{R}^{N_i \times N_i}, \quad i = 1, 2, \quad \hat{M} := \text{diag}(\hat{M}_1, \hat{M}_2) \in \mathbb{R}^{N \times N}. \]

Then, the complete model that takes into account the joint effect of the slow population growth in each patch and the slow and fast migrations between patches is given by
\[ dX(t) = \left( \frac{1}{\varepsilon} M + L + R \right) X(t) \, dt + \sum_{h=1}^{k} G^h X(t) \, dW^h(t) \]
or, using the fast time \( \tau \),
\[ dX(\tau) = (M + \varepsilon (L + R)) X(\tau) \, d\tau + \sqrt{\varepsilon} \sum_{h=1}^{k} G^h X(\tau) \, dW^h(\tau), \quad X(0) = X_0, \quad (14) \]
which constitutes a system of \( N \) linear SDEs.

Note that under the above conditions we are in the general setting of Section 3 by taking
\[ S = L + R, \quad F = M, \quad \hat{F} = \hat{M}, \quad B^h = G^h, \quad h = 1, \ldots, k \]
and Hypotheses S1, S2, F1, F2 and F3 for the slow and fast processes are met.

### 5.2. Model reduction

Therefore we can proceed to the aggregation of the original system following the procedure developed in Section 4. We define the global variables \( Y := UX \), that is,
\[ y^i(t) = 1_i^T x^i(t) = x_{i1}(t) + x_{i2}(t) + \cdots + x_{iN_i}(t), \quad i = 1, 2, \]
so that \( y^i \) is the total population in all patches of group \( i \), and the reduced aggregated system is the two dimensional SDE
\[ dY(t) = \tilde{S}Y(t) \, dt + \sum_{h=1}^{k} \tilde{G}^h Y(t) \, dW^h(t), \quad Y(0) = UX_0, \quad (15) \]
where
\[ \tilde{S} := U(L + R)V = \begin{pmatrix} 1_1^T (L_{11} + R_1)v_1 & 1_1^T L_{12} v_2 \\ 1_2^T L_{21} v_1 & 1_2^T (L_{22} + R_2)v_2 \end{pmatrix} \in \mathbb{R}^{2 \times 2} \]
\[ \tilde{G}^h := U G^h V = \text{diag}(1_1^T G^h v_1, 1_2^T G^h v_2), \quad h = 1, \ldots, k. \]

In order to be more specific we will consider the particular case in which \( N_1 = N_2 = 2 \), that is, there are 4 patches \((1, 1), (1, 2), (2, 1)\) and \((2, 2)\), and the population vector is \( X = (x_{11}, x_{12}, x_{21}, x_{22})^T \). Dispersal is fast between patches \((1, 1)\) and \((1, 2)\) and between \((2, 1)\) and \((2, 2)\), and is slow in the rest of cases.
Let $i = 1, 2$ be fixed. We will keep the notation introduced above except in the case of the fast migrations, where in order to simplify it we will denote by $p_i$ and $q_i$ the (fast) migration rates from $(i, 1)$ to $(i, 2)$ and from $(i, 2)$ to $(i, 1)$ respectively. Therefore

$$M_i = \begin{pmatrix} -p_i & q_i \\ p_i & -q_i \end{pmatrix}, \quad i = 1, 2.$$  

We assume $0 < p_i, q_i < 1$ so that matrix $M_i$ is irreducible and vector $v_i$ has the form $v_i = (q_i, p_i)^T / (p_i + q_i)$. Then the complete model has the form (14) with $G^h = \text{diag}(\sigma^h_{11}, \sigma^h_{12}, \sigma^h_{21}, \sigma^h_{22})$, $h = 1, \ldots, k$ and $M + \varepsilon (L + R)$ given by

$$
\begin{pmatrix}
-p_1 - \varepsilon(l_{12}^1 + l_{21}^1 - r_1^1) & q_1 & \varepsilon l_{12}^1 \\
q_1 & -p_1 - \varepsilon(l_{12}^2 + l_{21}^2 - r_1^2) & \varepsilon l_{12}^2 \\
\varepsilon l_{21}^1 & \varepsilon l_{21}^2 & -p_2 - \varepsilon(l_{12}^1 + l_{21}^1 - r_2^1) & q_2 \\
\varepsilon l_{21}^1 & \varepsilon l_{21}^2 & q_2 & -p_2 - \varepsilon(l_{12}^2 + l_{21}^2 - r_2^2)
\end{pmatrix}.
$$

The global variables are $Y = (y^1, y^2)^T = (x^{11} + x^{12}, x^{21} + x^{22})^T$ and the reduced system has the form (15) with

$$\tilde{S} = \begin{pmatrix} \tilde{s}_{11} \\ \tilde{s}_{21} \\ \tilde{s}_{22} \end{pmatrix},
$$

$$
\begin{pmatrix}
q_1(l_{11}^1 + l_{21}^1 - r_1^1) + p_1(-l_{11}^1 - l_{21}^1 + r_1^1) & q_1(l_{11}^2 + l_{21}^2 - r_1^2) + p_1(l_{12}^1 + l_{21}^2 - r_1^2) \\
q_1(l_{11}^1 + l_{21}^1 + p_1(-l_{11}^1 - l_{21}^1 + r_1^1) & q_1(l_{11}^2 + l_{21}^2 + p_1(l_{12}^1 + l_{21}^2 - r_1^2)
\end{pmatrix}.
$$

$$
\hat{G}^h = \text{diag}(\hat{\sigma}^h_1, \hat{\sigma}^h_2) = \text{diag}
\begin{pmatrix}
\frac{q_1l_{11}^1 + p_1l_{12}^1}{p_1 + q_1}, & \frac{q_2l_{12}^2 + p_2l_{22}^2}{p_2 + q_2}
\end{pmatrix}, \quad h = 1, \ldots, k.
$$

5.3. Case in which migration is fast

Let us now consider the particular but relevant case in which there is only one group of patches, and therefore all the migrations among the patches are fast with respect to demography. This has been the usual setting in the literature of approximate aggregation techniques [2, 3]. We can consider this case as a particular instance of the setting in Section 5.1 when we take $q$ equal to 1 instead of equal to 2. In order to simplify the notation, we omit the index 1 regarding the group of patches. Therefore, we denote the patches as $j = 1, \ldots, N$, and the vectors and matrices have the form $X = (x^1, \ldots, x^N)^T$, $R := \text{diag}(r^1, \ldots, r^N)$, $L = 0$ (there are no slow migrations), $M := [m^{ab}]$ and $G^h := \text{diag}(\sigma^h_1, \ldots, \sigma^h_N)$. Therefore, the original system (14) takes the form

$$
\text{d}X(\tau) = (M + \varepsilon R)X(\tau) \, d\tau + \sqrt{\varepsilon} \sum_{h=1}^k G^h X(\tau) \, dW^h(\tau).
$$

In this case there is only one global variable $y(t) = x^1(t) + x^2(t) + \cdots + x^N(t)$ and if $v = (v^1, \ldots, v^N)^T$ and $I = (1, \ldots, 1)^T$ are the positive right and left eigenvectors of $M$
associated to eigenvalue 0 scaled so that $1^T \mathbf{v} = 1$, we have $\mathbf{V} = \mathbf{v}$, $\mathbf{U} = 1^T$. Therefore, the reduced system is the scalar SDE

$$dy(t) = \bar{\mathbf{s}} y(t) \, dt + \sum_{h=1}^{k} \bar{g}^h y(t) \, dW^h(t), \quad y(0) = \sum_{j=1}^{N} \chi^j(0),$$

where

$$\bar{\mathbf{s}} := 1^T \mathbf{R} \mathbf{v} = \sum_{j=1}^{N} \mathbf{u}^j \mathbf{r}^j, \quad \bar{g}^h := 1^T \mathbf{G}^h \mathbf{v} = \sum_{j=1}^{N} \mathbf{u}^j \mathbf{\sigma}_j^h. \quad (17)$$

### 6. Relationships between the original and the reduced systems

The aim of this Section is to relate the asymptotic behaviour of the first- and second-order moments of the solution of the original and the reduced systems (8) and (10) introduced in Section 3. The two systems fit in the framework of [14] in which the quotient between the rate of diffusion squared and the speed of drift is $O(\varepsilon^p)$, $\varepsilon \to 0$, with $p > 1/2$, but the results obtained in that reference, essentially a stochastic extension of the Tikhonov theory of singular perturbations, are valid only for finite time intervals.

The first- and second-order moments of the solution of general linear stochastic equations with a deterministic initial condition are finite and can be calculated as the solution of certain linear ordinary differential equations. Specifically [1], for system (1) $\mathbb{E} \mathbf{X}(t)$ is the unique solution to the equation

$$\frac{d}{dt} \mathbb{E} \mathbf{X}(t) = \mathbf{A} \mathbb{E} \mathbf{X}(t), \quad \mathbb{E} \mathbf{X}(0) = \mathbf{X}_0,$$

whereas the matrix of second-order moments $\mathbf{H}(t) := \mathbb{E}(\mathbf{X}(t)\mathbf{X}(t)^T) \in \mathbb{R}^{d \times d}$ is the unique non-negative-definite symmetric solution of the equation

$$\frac{d \mathbf{H}(t)}{dt} = \mathbf{A} \mathbf{H}(t) + \mathbf{H}(t) \mathbf{A}^T + \sum_{h=1}^{k} \mathbf{B}^h \mathbf{H}(t) (\mathbf{B}^h)^T, \quad \mathbf{H}(0) = \mathbf{X}_0 \mathbf{X}_0^T. \quad (18)$$

Regarding Equation (18), in order to work with a more tractable expression, we will make use of the Kronecker matrix product and the ‘vec’ operator (see [9] for details). The Kronecker matrix product for two matrices $\mathbf{A} = [a_{ij}] \in \mathbb{C}^{m \times n}$ and $\mathbf{B} = [b_{ij}] \in \mathbb{C}^{r \times s}$ is defined as a matrix of size $mr \times ns$ with $mn$ blocks in which the block in position $(i,j)$ has the form $a_{ij} \mathbf{B}$. For any matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$, vec$\mathbf{A}$ is defined as the column vector that contains, in order, the columns of $\mathbf{A}$. For any matrices $\mathbf{A}$, $\mathbf{Y}$, and $\mathbf{B}$ for which the product $\mathbf{AYB}$ makes sense, one has vec($\mathbf{AYB}$) = ($\mathbf{B}^T \otimes \mathbf{A}$)vec$\mathbf{Y}$ [9, Theorem 6.8], and so applying the ‘vec’ operator to both sides of Equation (18) we obtain that the second-order moments verify

$$\frac{d}{dt} \text{vec} \mathbf{H}(t) = \mathbf{A}_2 \text{vec} \mathbf{H}(t), \quad \text{vec} \mathbf{H}(0) = \text{vec} (\mathbf{X}_0 \mathbf{X}_0^T),$$
where
\[
A_2 = I_N \otimes A + A \otimes I_N + \sum_{h=1}^{k} B^h \otimes B^h \in \mathbb{R}^{d^2 \times d^2}
\]
and \(I_N\) denotes the \(N \times N\) identity matrix.

Therefore, the asymptotic behaviour of the first- and second-order moments of system (1) can be characterized in terms of the dominant eigenvalue and eigenvectors of matrices \(A\) and \(A_2\). We will now use this fact to study the moments of systems (8) and (10).

In the case of the original system (8), the analogous to matrices \(A\) and \(A_2\) in the previous reasonings are matrices \(C \in \mathbb{R}^{N \times N}\) and \(C_2 \in \mathbb{R}^{N^2 \times N^2}\) given by
\[
C(\varepsilon) := F + \varepsilon S, \quad C_2(\varepsilon) := F_2 + \varepsilon S_2,
\]
where we have defined
\[
F_2 := I_N \otimes F + F \otimes I_N, \quad S_2 := I_N \otimes S + S \otimes I_N + \sum_{h=1}^{k} B^h \otimes B^h.
\]

In the case of the aggregated system (10) the corresponding matrices are \(\bar{C}(\varepsilon) = \varepsilon \bar{S} \in \mathbb{R}^{q \times q}, \quad \bar{C}_2(\varepsilon) = \varepsilon \bar{S}_2 \in \mathbb{R}^{q^2 \times q^2}\) where
\[
\bar{S}_2 := I_q \otimes \bar{S} + \bar{S} \otimes I_q + \sum_{h=1}^{k} \bar{B}^h \otimes \bar{B}^h.
\]

We introduce the following two hypotheses:

**Hypothesis 6.1:** Matrix \(\bar{S}\) has a simple and strictly dominant real eigenvalue \(\mu\) associated to non-negative right and left eigenvectors \(r\) and \(l\), respectively.

**Hypothesis 6.2:** Matrix \(\bar{S}_2\) has a simple and strictly dominant real eigenvalue \(\mu_2\) associated to non-negative right and left eigenvectors \(r_2\) and \(l_2\), respectively.

From Hypothesis 6.1 we have that the first-order moments of the reduced system have the following asymptotic behaviour
\[
\lim_{t \to \infty} \frac{\mathbb{E}Y(t)}{\exp(\mu t)} = \frac{\langle l, y_0 \rangle}{\langle l, r \rangle} r,
\]
where \(y_0 := UX_0\) and \(\langle \ast, \ast \rangle\) denotes the standard vector scalar product. Analogously, from Hypothesis 6.2 we have, for the second-order moments,
\[
\lim_{t \to \infty} \frac{\text{vec}\mathbb{E}(Y(t)Y(t)^T)}{\exp(\mu_2 t)} = \frac{\langle l_2, y_0 \otimes y_0 \rangle}{\langle l_2, r_2 \rangle} r_2.
\]

The next result relates the asymptotic behaviour of the first- and second-order moments of the original system (8) with that of the reduced system when \(\varepsilon\) is small, that is, when the separation between the time scales of migration and of demography is large enough:
Theorem 6.1: (a) Let Hypothesis 6.1 hold. Then for small enough \( \epsilon > 0 \), matrix \( C(\epsilon) = F + \epsilon S \) has a simple and strictly dominant eigenvalue \( \lambda(\epsilon) \) that can be written in the form
\[
\lambda(\epsilon) = \epsilon \mu + O(\epsilon^2)
\]
with associated right and left non-negative eigenvectors \( r(\epsilon) \) and \( l(\epsilon) \) such that
\[
r(\epsilon) = Vr + O(\epsilon); \quad l(\epsilon) = U^T l + O(\epsilon).
\]
Consequently, for the original system (8) we have:
\[
\lim_{\tau \to \infty} \frac{E X(\tau)}{\exp(\epsilon \mu + O(\epsilon^2))} = \langle l, UX_0 \rangle Vr + O(\epsilon).
\]
(20)

(b) Let Hypothesis 6.2 hold. Then for small enough \( \epsilon > 0 \), matrix \( C_2(\epsilon) = F_2 + \epsilon S_2 \) has a simple and strictly dominant eigenvalue \( \lambda_2(\epsilon) \) that can be written in the form
\[
\lambda_2(\epsilon) = \epsilon \mu_2 + O(\epsilon^2)
\]
with associated right and left non-negative eigenvectors \( r_2(\epsilon) \) and \( l_2(\epsilon) \) such that
\[
r_2(\epsilon) = (V \otimes V)r_2 + O(\epsilon); \quad l_2(\epsilon) = (U \otimes U)^T l_2 + O(\epsilon).
\]
(22)

Consequently, the asymptotic behaviour of the second-order moment of the original system is characterized by:
\[
\lim_{\tau \to \infty} \frac{\text{vec} E(X(t)X(t)^T)}{\exp(\epsilon \mu_2 + O(\epsilon^2))} = \langle l_2, (U \otimes U)\text{vec}(X_0X_0^T) \rangle (V \otimes V)r_2 + O(\epsilon).
\]
(23)

Proof: See Appendix.

From the last theorem it follows in particular that if \( \mu < 0 \) (resp. \( \mu > 0 \)), then \( \lambda(\epsilon) < 0 \) (resp. \( \lambda(\epsilon) > 0 \)) for small enough \( \epsilon \), so that if the expected value of the population vector tends to zero (resp. infinity) in the reduced system the same happens for the original one when \( \epsilon \) is small. Something analogous happens for the second-order moments regarding \( \mu_2 \) and \( \lambda_2(\epsilon) \).

Let us apply this result to relate the behaviour of the multiregional models of Section 5. In the first place we will consider the model with \( q = 2, N_1 = N_2 = 2 \). Note from (16) that if there is at least a non-zero coefficient \( l_{ij}^\beta \), matrix \( \bar{S} \) is irreducible and then Hypothesis 6.1 holds. Its dominant eigenvalue is given by
\[
\mu = \frac{\bar{s}_{11} + \bar{s}_{22} + \sqrt{(\bar{s}_{11} - \bar{s}_{22})^2 + 4\bar{s}_{12}\bar{s}_{21}}}{2}
\]
and so the rate of growth of the first-order moments for the original system is \( \epsilon \mu + O(\epsilon^2) \) for small enough \( \epsilon \). This expression allows one to study how the different combinations of the migration and growth parameters affect the asymptotic behaviour of the expected value of the population. In particular, \( \mu < 0 \) if and only if \( \text{tr} \bar{S} < 0 \) and \( \det \bar{S} > 0 \). Since \( \bar{S} \) has order two we can calculate right and left eigenvectors \( r \) and \( l \) associated to \( \mu \) and
then apply Equations (20) and (23) to obtain the full information about the asymptotic behavior of the first-order moments of the system.

We could argue analogously for the second-order moments: in this case matrix $\bar{S}_2$ is of order four and so an analytical computation of $\mu_2$, $r_2$ and $l_2$ is non-feasible, but for instance we can use the Routh–Hurwitz criterion [17] to obtain conditions for $\mu_2$ to be negative, so that $\varepsilon \mu_2 + O(\varepsilon^2)$ will also be negative for small $\varepsilon$.

In the case of the system of Section 5.3, the analysis is simpler cause the aggregated system is scalar and so are $\bar{S}$ and $\bar{S}_2$, so that Hypotheses 6.1 and 6.2 hold trivially. Moreover, $\mu = \bar{s}$ and $\mu_2 = 2\bar{s} + \sum_{h=1}^{k} (g^h)^2$ where $\bar{s}$ and the $g^h$ are given by Equation (17) and we can take $r = l = r_2 = l_2 = 1$. Therefore from Equations (20) and (23) we have

$$\lim_{\tau \to \infty} \frac{E(X(\tau))}{\exp(\varepsilon \mu + O(\varepsilon^2))} = y(0)v + O(\varepsilon),$$

$$\lim_{\tau \to \infty} \text{vec}E(X(t)X(t)^T)\exp(\varepsilon \mu_2 + O(\varepsilon^2)) = y(0)^2(v \otimes v) + O(\varepsilon).$$

Note in particular that, when $\varepsilon$ is small enough, the expected value of the population vector tends to zero if $\bar{s} < 0$, and the second-order moments also tend to zero if $\bar{s} < -\sum_{h=1}^{k} (g^h)^2 / 2$.

7. Discussion

In this work we have presented a technique to carry out the reduction of linear two-scale population models governed by SDEs, therefore allowing one to simplify the treatment of these models. The reduction of the original model with $N$ variables is carried out by letting the fast process reach equilibrium and defining an appropriate set of $q$ global variables, $q < N$, which are linear combinations of the state variables and are conservative for the fast process. Moreover, we have obtained conditions that guarantee the positivity of solutions to the model.

We have also presented a result that allows one to know the the asymptotic behaviour of the first- and second-order statistical moments of the population vector of the original system through the computation of the dominant eigenvalues and eigenvectors of certain matrices of dimensions $q \times q$ and $q^2 \times q^2$ associated to the reduced system.

The aggregation technique has been applied to study stochastic multiregional models in which migration between some patches can be fast with respect to demography whereas dispersal between other patches happens at the scale of demography. In the simpler case in which all migrations are fast, the reduced system is a scalar SDE whose analysis is straightforward, therefore allowing one to easily characterize the asymptotic behaviour of those moments for the original multiregional model.

This work suggests possible lines of future development: on the first hand, and still within the linear setting, work needs to be done to try to relate the stochastic stability [15] of the origin in the original and the reduced system, for this will provide more information regarding the persistence-extinction of the population. Secondly, in order to be able to study more realistic population models, the technique should be extended to nonlinear settings.
Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
Authors are supported by MINECO (Ministerio de Economía, Industria y Competitividad), Spain, project MTM2014-56022-C2-1-P.

ORCID
Luis Sanz http://orcid.org/0000-0002-1054-4568

References
[1] L. Arnold, Stochastic Differential Equations: Theory and Applications, Wiley Interscience, New York, 1974.
[2] P. Auger, R.B. de La Parra, J.-C. Poggiale, E. Sánchez, and L. Sanz, Aggregation methods in dynamical systems and applications in population and community dynamics, Phys. Life Rev. 5(2) (2008), pp. 79–105.
[3] P. Auger, J. Poggiale, and E. Sánchez, A review on spatial aggregation methods involving several time scales, Ecol. Complex. 10 (2012), pp. 12–25.
[4] H. Baumgärtel, Analytic Perturbation Theory for Matrices and Operators, Vol. 15, Springer, New York, 1985.
[5] N. Berglund and B. Gentz, Geometric singular perturbation theory for stochastic differential equations, J. Differential Equations 191(1) (2003), pp. 1–54.
[6] C.A. Braumann, Variable effort fishing models in random environments, Math. Biosci. 156(1) (1999), pp. 1–19.
[7] C.A. Braumann, Harvesting in a random environment: Itô or stratonovich calculus? J. Theoret. Biol. 244(3) (2007), pp. 424–432.
[8] C. Carlos and C.A. Braumann, General population growth models with allele effects in a random environment, Ecol. Complex. 30 (2016), pp. 26–33.
[9] M. Fiedler, Special Matrices and their Applications in Numerical Mathematics, Kluwer Boston, Inc., 1986.
[10] N. Herath and D. Del Vecchio, Model order reduction for linear noise approximation using time-scale separation, in 2016 IEEE 55th Conference on Decision and Control (CDC), IEEE, Chicago, IL, USA, 2016, pp. 5875–5880.
[11] N. Herath and D. Del Vecchio, Model reduction for a class of singularly perturbed stochastic differential equations: Fast variable approximation, in American Control Conference (ACC), 2016, IEEE, Boston, MA, USA, 2016, pp. 3674–3679.
[12] R.A. Horn and C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 2012.
[13] D. Jiang, J. Yu, C. Ji, and N. Shi, Asymptotic behavior of global positive solution to a stochastic sir model, Math. Comput. Modelling 54(1) (2011), pp. 221–232.
[14] Y. Kabanov and S. Pergamenshchikov, Two-Scale Stochastic Systems: Asymptotic Analysis and Control, Vol. 49, Springer, Berlin, 2003.
[15] R. Khasminskii, Stochastic Stability of Differential Equations, Vol. 66, Springer, Berlin, 2012.
[16] H.J. Kushner, Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems, Vol. 3, Birkhäuser, Boston, 1990.
[17] J.A. Linda, An Introduction to Mathematical Biology, Pearson, Upper Saddle River, NJ, 2007.
[18] X. Mao, Stochastic Differential Equations and their Applications, Woodhead Publishing; 2nd ed. (January 13, 2008), Oxford, 1997.
[19] L. Sanz and J. Alonso, Approximate aggregation methods in discrete time stochastic population models, Math. Model. Nat. Pheno. 5(6) (2010), pp. 38–69.
Appendix

Proof of Theorem 2.1. Let us fix $X_0 > 0$ and let $k_0$ be sufficiently large so that for every $i = 1, \ldots, d$, the $i$th component of $X_0$ verifies $X_0^i > 1/k_0$. For each integer $k \geq k_0$ we define the stopping time $\tau_k := \inf \{ t \geq 0 : X^i(t) \leq 1/k \text{ for some } i = 1, \ldots, d \}$. Clearly $\tau_k$ is increasing as $k \to \infty$. Let us set $\tau_\infty := \lim_{k \to \infty} \tau_k$. We want to show that $\tau_\infty = \infty$ with probability one, and to do so we will proceed by contradiction. Let us assume that the previous statement is false. Then there exists $T > 0$ and $\epsilon \in (0, 1)$ such that $P(\tau_\infty \leq T) > \epsilon$, and therefore there exists an integer $k_1 \geq k_0$ such that

$$P(\tau_k \leq T) \geq \epsilon \quad \text{for all } k \geq k_1. \quad (A1)$$

Let $R^d_+ = (0, \infty) \times \cdots \times (0, \infty)$, let $v : R^d_+ \to R$ be any $C^2$ function and let $V(t) := v(X(t))$. From Itô’s theorem [1] we have

$$dV(t) = Lv(X(t)) \, dt + \frac{\partial v}{\partial x}(X(t)) \sum_{h=1}^m B^h X(t) \, dW^h(t), \quad (A2)$$

where

$$Lv(x) = \sum_{i,j=1}^d \frac{\partial v}{\partial x_i} a_{ij} x_j + \frac{1}{2} \sum_{h=1}^m \sum_{i,j=1}^d (B^h x x^T (B^h)^T)_{ij} x_i x_j \frac{\partial^2 v}{\partial x_i \partial x_j}$$

$$= \sum_{i,j=1}^d \frac{\partial v}{\partial x_i} a_{ij} x_j + \frac{1}{2} \sum_{h=1}^m \sum_{i,j=1}^d b_{ij}^h x_i x_j \frac{\partial^2 v}{\partial x_i \partial x_j}$$

and we have used that the $B^h$ are diagonal, $B^h = \text{diag}(b_{11}^h, \ldots, b_{dd}^h)$. Let us choose $v(x) := -\sum_{i=1}^d \log x_i / (1 + x_i)$, which is positive and $C^2(R^d_+)$. Straightforward calculations show that

$$Lv(x) = -\sum_{i=1}^d \frac{a_{ii}}{1 + x_i} - \sum_{i=1}^d \sum_{j \neq i} \frac{a_{ij} x_j}{1 + x_i} + \frac{1}{2} \sum_{h=1}^m \sum_{i=1}^d (b_{ii}^h)^2 \frac{1 + 2x_i}{(1 + x_i)^2}. \quad (A3)$$

Note that, if $x_i > 0$ for all $i = 1, \ldots, d$, then the first and the third term in Equation (A3) are bounded from above in $R^d_+$, and moreover, since $A$ is a Metzler matrix, the second term is negative in $R^d_+$. Therefore there exists $K \geq 0$ such that $Lv(x) \leq K$ in $R^d_+$. Clearly $X(t \wedge \tau_k)$, where we define $a \wedge b := \min(a, b)$, belongs to $R^d_+$ for all $t \geq 0$, and therefore, using the previous bound we have from Equation (A2)

$$V(X(T \wedge \tau_k)) \leq V(X_0) + K(T \wedge \tau_k) + \int_0^{T \wedge \tau_k} \frac{\partial v}{\partial x}(X(t)) \sum_{h=1}^m B^h X(t) \, dW^h(t)$$

and therefore, using that the expected value of the stochastic integral is zero

$$\mathbb{E} V(X(T \wedge \tau_k)) \leq V(X_0) + K\mathbb{E}(T \wedge \tau_k) \leq V(X(0)) + KT. \quad (A4)$$
Let \( \mu(\epsilon) \) be a perturbed eigenvalue such that \( \lambda(\epsilon) \neq 0 \), \( \mu(\epsilon) \) is simple for \( t = (\epsilon) \), it follows that

\[
\lambda(\epsilon) = \lambda_0 + \mu_1 \epsilon + o(\epsilon), \quad \epsilon \to 0,
\]

for some \( j = 1, \ldots, q \), where \( \mu_j, j = 1, \ldots, q \) are the eigenvalues of the restriction \( \mathbf{T} \mid_{\text{Im} \mathbf{P}} \) of operator \( \mathbf{T} := \mathbf{PA}_1 \mathbf{P} \) to \( \text{Im} \mathbf{P} \). Moreover, associated to \( \lambda(\epsilon) \) we can choose a (right) eigenvector \( \mathbf{x}(\epsilon) \) that is continuous at \( \epsilon = 0 \) and such that \( \mathbf{x}_0 := \mathbf{x}(0) \neq 0 \) verifies

\[
\mathbf{T} \mathbf{x}_0 = \mu_0 \mathbf{x}_0.
\]

(2) In the case that \( \mu \) is a simple eigenvalue of \( \mathbf{T} \), \( \lambda(\epsilon) \) and \( \mathbf{x}(\epsilon) \) are holomorphic functions of \( \epsilon \).

**Proof of Theorem 6.1:** For small \( \epsilon \) we will consider matrix \( \mathbf{C}(\epsilon) = \mathbf{F} + \epsilon \mathbf{S} \) as a perturbation of matrix \( \mathbf{F} \). From Equation (4) and the properties of each matrix \( \mathbf{F}_i \), we have that \( 0 \) is a semisimple eigenvalue of \( \mathbf{F} \) and it is strictly dominant. Therefore

\[
\mathbf{F} = (\mathbf{V} \mid \mathbf{V}')\text{diag}(0_q, \Sigma') \begin{pmatrix} \mathbf{U} \\ \mathbf{U}' \end{pmatrix} \tag{A5}
\]

is a Jordan canonical decomposition of \( \mathbf{F} \) [12] where \( \mathbf{V}' \in \mathbb{R}^{n \times (N-q)} \), \( \mathbf{U}' \in \mathbb{R}^{n \times (N-q)} \) and \( \Sigma' \in \mathbb{R}^{(N-q) \times (N-q)} \) are certain matrices and \( \Sigma' \) is associated to eigenvalues with strictly negative real part. From Equation (A5) and Lemma 4.1, we have that \( \tilde{\mathbf{F}} \equiv \mathbf{VU} \) is the eigenprojection matrix of \( \mathbf{F} \) associated to eigenvalue \( 0 \).

Using the continuity of eigenvalues on matrix entries, the dominant eigenvalue of \( \mathbf{C}(\epsilon) \) for small enough \( \epsilon \) necessarily must belong to the set of eigenvalues of \( \mathbf{C}(\epsilon) \) that tend to \( 0 \) when \( \epsilon \to 0 \). Using part 1 of Theorem A.1 applied to \( \mathbf{C}(\epsilon) = \mathbf{F} + \epsilon \mathbf{S} \) with \( \mathbf{A}_0 = \mathbf{F}, \mathbf{A}_1 = \mathbf{S}, \lambda_0 = 0 \) and \( \mathbf{P} = \tilde{\mathbf{F}} \), we conclude that this set of eigenvalues has the form

\[
0 + \epsilon \mu_j + o(\epsilon), \quad j = 1, \ldots, q, \tag{A6}
\]

where the \( \mu_j \) are the eigenvalues of the restriction \( \mathbf{T} \mid_{\text{Im} \tilde{\mathbf{F}}} \) of \( \tilde{\mathbf{F}} \tilde{\mathbf{S}} \tilde{\mathbf{F}} \) to \( \text{Im} \tilde{\mathbf{F}} \). Clearly \( \text{Im} \tilde{\mathbf{F}} \) is invariant for \( \mathbf{T} \) and, using Equation (11) and Lemma 4.1, we have that \( \mathbf{T} = \tilde{\mathbf{F}} \tilde{\mathbf{S}} \tilde{\mathbf{F}} = \mathbf{VUSVU} = \mathbf{VSU} \), so that \( \tilde{\mathbf{S}} \) is the matrix of the restriction \( \mathbf{T} \mid_{\text{Im} \tilde{\mathbf{F}}} \) expressed in the basis of \( \text{Im} \tilde{\mathbf{F}} \) defined by the columns of \( \mathbf{V} \). Therefore, \( \mathbf{T} \mid_{\text{Im} \tilde{\mathbf{F}}} \) and \( \tilde{\mathbf{S}} \) have the same eigenvalues and, moreover, if \( \mathbf{x} \) and \( \mathbf{y} \) are, respectively, right and left eigenvectors of \( \tilde{\mathbf{S}} \) associated to a certain eigenvalue \( \lambda \), then \( \mathbf{Vx} \neq 0 \) and \( \mathbf{U}^\top \mathbf{y} \neq 0 \) are right and left eigenvectors of \( \mathbf{T} \mid_{\text{Im} \tilde{\mathbf{F}}} \) associated to eigenvalue \( \lambda \).

Therefore, using Hypothesis 6.1 it follows that \( \mu \) is a simple eigenvalue for \( \mathbf{T} \mid_{\text{Im} \tilde{\mathbf{F}}} \) and is associated to right and left eigenvectors \( \mathbf{Vr} \neq 0 \) and \( \mathbf{U}^\top \mathbf{r} \neq 0 \), respectively. Moreover, \( \mu \) is strictly dominant for \( \mathbf{T} \mid_{\text{Im} \tilde{\mathbf{F}}} \) so that, using Equation (A6), \( \mu(\epsilon) := \epsilon \mu + o(\epsilon) \) is a simple eigenvalue for \( \mathbf{C}(\epsilon) \) and strictly dominant when \( \epsilon \) is small enough.

Now, using part 2 of Theorem A.1 and the fact that \( \mu \) is simple for \( \mathbf{T} \mid_{\text{Im} \tilde{\mathbf{F}}} \), it follows that \( \mu(\epsilon) \) has the form \( \mu(\epsilon) := \epsilon \mu + O(\epsilon^2) \) and that is associated to right and left eigenvectors which can be
written in the form $Vr + O(\varepsilon)$ and $U^T I + O(\varepsilon)$ respectively. Then, for small enough $\varepsilon$

$$\lim_{\tau \to \infty} \frac{\mathbb{E}X(\tau)}{\exp(\varepsilon \mu + O(\varepsilon^2))} = \frac{(I^T U + O(\varepsilon))X_0}{(I^T U + O(\varepsilon))(Vr + O(\varepsilon))} (Vr + O(\varepsilon)) = \frac{(I, UX_0)}{(I, r)} Vr + O(\varepsilon)$$

and so Equation (20) is proved.

(b) The proof is analogous to that of part (a). A notable property of the Kronecker product that we will use in the sequel is [9, Theorem 6.1]

$$(A_1 \otimes A_2 \otimes \cdots \otimes A_m)(B_1 \otimes B_2 \otimes \cdots \otimes B_m) = (A_1 B_1) \otimes (A_2 B_2) \otimes \cdots \otimes (A_m B_m).$$

We will use Equation (19) and for small $\varepsilon$ will consider matrix $C(\varepsilon)$ as a perturbation of matrix $F_2 = I_N \otimes F + F \otimes I_N$. First, let us study the spectral properties of $F_2$. Let $\hat{v}_i, i = 1, \ldots, N$, denote the $i$th column of matrix $(V | V')$, that is, $\hat{v}_i$ is an eigenvector of $F$ associated to eigenvalue 0 if $1 \leq i \leq q$ or to an eigenvalue with strictly negative real part if $q + 1 \leq i \leq N$.

From [9, Theorem 6.5] we know that the eigenvalues of $I_N \otimes F + F \otimes I_N$ are the set $\{\lambda_i + \lambda_j : i = 1, \ldots, N\}$ and that, associated to each eigenvalue $\lambda_i + \lambda_j$ we have right and left eigenvectors $\hat{v}_i \otimes \hat{v}_j$ and $\hat{u}_i \otimes \hat{u}_j$, where $\hat{v}_i$ and $\hat{u}_i$ are, respectively, right and left eigenvectors of $F$ associated to eigenvalue $\lambda_i$, $i = 1, \ldots, N$. Taking into account the spectrum of $F$ we conclude that $\sigma(F_2)$ consists of $q$ eigenvalues with multiplicity $q^2$ associated to right (resp. left) eigenvectors $\hat{v}_i \otimes \hat{v}_j$, (resp. $\hat{u}_i \otimes \hat{u}_j$), where either $i$ or $j$ belong to the set $\{q + 1, \ldots, N\}$. Clearly, eigenvalue $\lambda = 0$ is semisimple and is strictly dominant for $F_2$.

It is easy to check that if $A$ and $B$ are any matrices with the block structure $A = (a_1 | \cdots | a_k)$, $B = (B_1 | \cdots | B_s)$ where the $a_i$ are column vectors, then

$$A \otimes B = (a_1 \otimes B_1 | \cdots | a_1 \otimes B_s | \cdots | a_k \otimes B_1 | \cdots | a_k \otimes B_s).$$

Applying this result we can write

$$(\hat{v}_1 \otimes \hat{v}_1 | \cdots | \hat{v}_1 \otimes \hat{v}_q | \cdots | \hat{v}_q \otimes \hat{v}_1 | \cdots | \hat{v}_q \otimes \hat{v}_q) = V \otimes V$$

$$(\hat{u}_1 \otimes \hat{u}_1 | \cdots | \hat{u}_1 \otimes \hat{u}_q | \cdots | \hat{u}_q \otimes \hat{u}_1 | \cdots | \hat{u}_q \otimes \hat{u}_q) = U^T \otimes U^T$$

so that we can write a Jordan canonical decomposition of $F_2$ in the form

$$F_2 = (V \otimes V | V_2') \text{diag}(0, \Sigma_2') \left( \frac{U \otimes U}{U_2'} \right),$$

where $V_2'$, $U_2'$ and $\Sigma_2'$ are appropriate matrices. From this expression we see that $\hat{F}_2 := (V \otimes V)(U \otimes U) = (VV) \otimes (VV) = \hat{F} \otimes \hat{F}$ is the eigenprojection of $F_2$ corresponding to eigenvalue 0.

Like in part (a), we know that, for small enough $\varepsilon$, the dominant eigenvalue of $C_2(\varepsilon)$ necessarily must belong to the set of eigenvalues of $C_2(\varepsilon)$ that tend to 0 when $\varepsilon \to 0$. Let us consider the restriction $T_2 \mid_{\text{Im} \hat{F}_2}$ of matrix $T_2 := \hat{F}_2 S_2 \hat{F}_2$ to $\text{Im} \hat{F}_2$. Clearly $\text{Im} \hat{F}_2$ is invariant for $T_2$ and, using Equation (11), Lemma 4.1 and the properties of the Kronecker product we have $T_2 = \hat{F}_2 S_2 \hat{F}_2 = (V \otimes V)S_2(U \otimes U)$ so that $S_2$ is the matrix of $T_2 \mid_{\text{Im} \hat{F}_2}$ expressed in the basis of $\text{Im} \hat{F}_2$ defined by the columns of $V \otimes V$. Therefore, $T_2 \mid_{\text{Im} \hat{F}_2}$ and $S_2$ have the same eigenvalues and, moreover, if $x$ and $y$ are, respectively, right and left eigenvectors of $S_2$ associated to a certain eigenvalue $\lambda$, then $(V \otimes V)x \neq 0$ and $(U \otimes U)^T y \neq 0$ are right and left eigenvectors of $T_2 \mid_{\text{Im} \hat{F}_2}$ associated to eigenvalue $\lambda$.

Now let us assume Hypothesis 6.2. Then we can reason exactly as we did in part (a), substituting $C(\varepsilon)$, $F$, $S$, $\hat{F}$, $V$ and $U$ for $C_2(\varepsilon)$, $F_2$, $S_2$, $\hat{F}_2$, $V \otimes V$ and $U \otimes U$, respectively, and obtain that for small enough $\varepsilon$, $C_2(\varepsilon)$ has a simple and strictly dominant eigenvalue that can be written in the form (21) and associated to eigenvectors with the form (22), from where the asymptotic behaviour (23) follows.