Mathematical Statistics
Stockholm University

Quasi Equilibrium, Variance Effective Size and Fixation Index for Populations with Substructure

Ola Hössjer
Nils Ryman

Research Report 2012:4

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Research Report **2012:4**,
http://www.math.su.se/matstat

Quasi Equilibrium, Variance Effective Size and Fixation Index for Populations with Substructure

Ola Hössjer*
Nils Ryman†

February 2012

Abstract

We present a general method for joint computation of the variance effective size N_{eV}, the fixation index F_{ST} and the coefficient of gene differentiation G_{ST} for a class of structured populations with very general migration and reproduction properties, where the size of each particular subpopulation is constant in time. Our approach is relevant when migration rates are of larger magnitude than mutation rates, so that new mutations can be ignored before quasi equilibrium between genetic drift and migration is obtained. The vector valued time series of subpopulation allele frequencies is divided into two parts; one corresponding to genetic drift of the whole population and one corresponding to the dynamics of fluctuations of allele frequencies between subpopulations. After a simple standardization, the latter is shown to be in equilibrium, conditional on that no allele is fixed in the population. This enables us to compute quasi equilibrium approximations of N_{eV}, F_{ST} and G_{ST}.

Our findings are illustrated for several reproduction and migration scenarios, including the island model, stepping stone models and a model where one subpopulation acts as a genetic reservoir. We also discuss overlapping generations and extensions to randomly varying subpopulation sizes.

Key words: Autoregressive time series, island model, quasi equilibrium, stepping stone models, spatial allele frequency fluctuations, structured populations, temporal allele frequency fluctuations.
Mathematics Subject Classification (2000): 92D25, 60F99

---

*Department of Mathematics, Div. of Mathematical Statistics, Stockholm University.
†Department of Zoology, Div. of Population Genetics, Stockholm University.

1 Introduction

Most populations exhibit some degree of spatial, age, size, social or ethnic heterogeneity. This structure is often quite complicated and influences the dynamics of the population. From a genetic point of view, a few summary statistics of the population are often extracted that broadly characterize its main properties. The effective population size was introduced by Wright (1931, 1938). It is the size of an ideal population exhibiting the same rate of genetic drift as the studied one. For models with spatial substructure, the fixation index of Wright (1951) quantifies the degree of heterogeneity between subpopulations.

Many closely related definitions of the effective population size and fixation index exist, see for instance Orrive (1993), Caballero (1994), Wang and Caballero (1999), Waples (2002), Ewens (2004), Ryman and Leimar (2008) and Hössjer et al. (2012). We will focus on definitions in terms of variance of allele frequency differences, over time and space respectively. Once a mutation has occurred at a given locus, we will disregard the occurrence of *new* mutations at the same locus before equilibrium is attained. Consequently, in absence of new mutations, the effective size of the population quantifies the rate of loss of heterozygosity or rate of increased inbreeding. It is therefore important to assess this size and prevent it from getting too low, see Palstra and Ruzzante (2008) and Hare et al. (2011) for recent reviews for wildlife populations, in particular marine species.

Equilibrium values of the fixation index can be derived, for the infinite island model, from results of Sved and Latter (1977). For finite populations, no equilibrium value exists in absence of new mutations, since some allele will eventually get fixed at any locus. However, Hössjer et al. (2012) derived quasi equilibrium values of the fixation index for the finite island model, conditionally on no fixation. We now extend this approach to joint computation of the fixation index and effective population size at quasi equilibrium, for a large class of finite population models. This class includes general migration schemes between and reproduction schemes within subpopulations, subject only to the constraint of constant subpopulation sizes over time.

The paper is organized as follows: In Sections 2 and 3 we define the dynamics of how genes are passed on from one time point to the next in the context of one biallelic locus. In particular, we define the effective population size, the fixation index and the vector valued and time inhomogeneous process of subpopulation allele frequencies. In Section 4 we demonstrate how this process can be decomposed into two parts, one corresponding to genetic drift of the whole population and the other to fluctuations of allele frequencies between subpopulations. After a simple standardization, the latter becomes a (quasi) stationary process. Based on this, we present in Section 5 a general algorithm for computing the quasi equilibrium effective population size and fixation index. To illustrate the generality of our approach, several examples of reproduction and spatial migration are presented in Sections 6 and 7. In Section 8 we consider multiallelic and multilocus extensions and show that the coefficient of gene differentiation of Nei (1973) has the same quasi equilibrium value as the fixation index. Numerical results are presented in Section 9 and possible extensions are briefly discussed in Section 10. Finally, the more technical derivations are presented in the appendix.

2 Migration Model

Consider a population with N diploid individuals, each one having two copies of a particular genetic marker, i.e. a fixed portion of DNA. We will simply refer to this marker as a gene, so that the population consists of $2N$ genes. We further assume that the population can be divided into $s \geq 2$ subpopulations. These may refer to geographical sites, age classes, social or ethnic groups of combinations thereof, see for instance Nordborg and Krone (2002), Sagitov and Jagers (2005), Hössjer (2011) and references therein. Let a_i refer to the relative size of subpopulation i, so that

$$\sum_{i=1}^{s} a_i = 1. \tag{1}$$

We further assume that the population evolves in discrete time $t = 0, 1, \ldots$ with relative subpopulation sizes that remain constant. Hence the number of diploid individuals of subpopulation i is Na_i in any generation.

Consider a fixed generation t and number the genes of each subpopulation k as $l = 1, \ldots, 2Na_k$. In order to describe how genes of generation t are passed on to the next generation $t+1$, we introduce $\nu_{t,ki}^l = \nu_{ki}^l$ as the number of copies of a particular gene l of subpopulation k in generation t that are passed on to subpopulation i in the next generation. Passing on should here be interpreted broadly, either as being transmitted to offspring individuals, or that the individual carrying the gene survives to the next generation and migrates from k to i. In order to keep subpopulation sizes constant, we require

$$\sum_{k=1}^{s} \sum_{l=1}^{2Na_k} \nu_{ki}^l = 2Na_i, \quad i = 1, \ldots, s. \tag{2}$$

We will also introduce the migration rate m_{ki} between subpopulations k and i, by assuming

$$\sum_{l=1}^{2Na_k} \nu_{ki}^l = 2Na_k m_{ki}(1 + o(1)), \tag{3}$$

where $o(1)$ is a remainder term that tends to zero as N grows. If $\{\nu_{ki}^l\}_{l=1}^{2Na_k}$ are exchangeable random variables, it follows from (3) that

$$E(\nu_{ki}^l) = m_{ki}(1 + o(1)).$$

Hence, m_{ki} is essentially the average number of copies that the genes of subpopulation k pass on to subpopulation i in the next generation.

Combining (2) and (3), we find that

$$\sum_{k=1}^{s} a_k m_{ki} = a_i \tag{4}$$

for all $i = 1, \ldots, s$. This implies that $\boldsymbol{a} = (a_1, \ldots, a_s)$ is a left eigenvector of the migration matrix $\boldsymbol{M} = (m_{ki})$ with eigenvalue 1 (Caswell, 2001). Reversing time, the probability that the parent of a gene of subpopulation i originates from subpopulation k is of the order

$$b_{ik} = \frac{a_k m_{ki}}{a_i}. \tag{5}$$

Because of (4), it is easy to see that $\boldsymbol{B} = (b_{ik})$ is the transition matrix of a Markov chain, having row sums 1. This Markov chain is assumed to be irreducible and aperiodic, with asymptotic distribution $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_s)$. In general $\boldsymbol{\gamma}$ differs from $\boldsymbol{a}$, although they agree for many well known migration models, see Nagylaki (1980) and Subsection 7.1 for further details. The global migration rate

$$m' = 1 - \sum_{i=1}^{s} a_i m_{ii} = 1 - \sum_{i=1}^{s} a_i b_{ii} \tag{6}$$

is defined as the fraction of offspring of the whole population that in each generation migrates to another subpopulation. Complementary to m', it is also important to know whether migration in one step is local or global. To this end, we define the one step neighborhood $\mathcal{N}_k = \{i;\, i \neq k, m_{ki} > 0\}$ of each subpopulation k as well as their average relative size

$$|\mathcal{N}| = \frac{1}{s-1} \sum_{k=1}^{s} a_k |\mathcal{N}_k|, \tag{7}$$

which is a different concept than the neighborhood size NS of Wright (1946), developed for continuous spatial isolation by distance models.

Table 1 contains a list of the most important symbols used in the paper.

3 Effective Population Size and Fixation Index

Assuming a biallelic gene, we let P_{ti} denote the fraction of subpopulation i at generation t with a given allele, so that

$$p_t = \sum_{i=1}^{s} a_i P_{ti} \tag{8}$$

refers to the frequency of the same allele in the whole population at time t. In order to describe the time dynamics of the allele frequencies, we notice that $p_{t+1,i}$ would equal $\sum_{k=1}^{s} b_{ik} P_{tk}$ if individuals of subpopulation i received genes in exact proportions $b_{i1}, \ldots, b_{is}$ from the subpopulations of the parental generation, and in addition the allele frequencies of genetic material transmitted from subpopulation k to i was identical to P_{tk}, the allele frequency of the parental subpopulation k. However, the subpopulation proportions of parental origin as well as transmitted allele frequencies from the various subpopulations will exhibit some random variation, motivating the recursion formula

$$\boldsymbol{P}_{t+1} = \boldsymbol{B}\boldsymbol{P}_t + \boldsymbol{\varepsilon}_{t+1} \tag{9}$$

for the vector $\boldsymbol{P}_t = (P_{t1}, \ldots, P_{ts})^T$ of subpopulation allele frequencies, with T denoting matrix transposition. The random error term $\boldsymbol{\varepsilon}_{t+1}$ is assumed to satisfy

$$E(\boldsymbol{\varepsilon}_{t+1} | \boldsymbol{P}_t) = 0, \tag{10}$$

corresponding to a selectively neutral allele, with covariance matrix

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_{t+1} | \boldsymbol{P}_t) = \boldsymbol{\Sigma}(\boldsymbol{P}_t), \tag{11}$$

the form of which will depend on the reproduction model. Hence $\boldsymbol{P}_t$ is a multivariate autoregressive time series (Brockwell and Davis, 1987) with some degree of heteroscedasticity, since $\boldsymbol{\Sigma}(\boldsymbol{P}_t)$ depends on $\boldsymbol{P}_t$.

The two quantities of main interest for us are the variance effective population size

$$N_{eV,t} = \frac{p_t(1-p_t)}{2\mathrm{Var}(p_{t+1} - p_t | p_t)} \tag{12}$$

and fixation index

$$F_{ST,t} = \frac{\sum_{i=1}^{s} a_i (P_{ti} - p_t)^2}{p_t(1-p_t)}, \tag{13}$$

see for instance equation (12.13) in Nei and Kumar (2000).

4 Quasi Equilibrium

We will study the long run behaviour of $N_{eV,t}$ and $F_{ST,t}$ before one allele gets fixed in all subpopulations, so called quasi equilibrium. This means that conditionally no fixation, the joint distribution of $N_{eV,t}$ and $F_{ST,t}$ will converge, so that they exhibit fluctuations around means F_{ST}^{eq} and N_{eV}^{eq}, referred to as the quasi equilibrium values of the fixation index and effective population size.

Our findings will depend crucially on the properties of the vector valued time series of allele frequencies $\boldsymbol{P}_t$. Since $\boldsymbol{B}$ has largest eigenvalue 1, this time series is non-stationary, and therefore no equilibrium solution exists. However, for large enough populations, $\boldsymbol{P}_t$ can be decomposed into a sum of two parts;

$$\boldsymbol{P}_t = P_t\mathbf{1} + \boldsymbol{P}_t^0, \tag{14}$$

where $\mathbf{1} = (1, \ldots, 1)^T$ and

$$P_t = \sum_{i=1}^{s} \gamma_i P_{ti} = \boldsymbol{\gamma}\boldsymbol{P}_t. \tag{15}$$

The first term on the right hand side of (14) is non-stationary. It describes how the average allele frequency drifts with time. Subpopulation allele frequencies are weighted together according to $\boldsymbol{\gamma}$ rather than $\boldsymbol{a}$, and therefore P_t typically differs slightly from p_t, unless $\boldsymbol{a} = \boldsymbol{\gamma}$. The second term $\boldsymbol{P}_t^0$ of (14) is also non-stationary, and it corresponds to local subpopulation allele frequency fluctuations around P_t. However, we will see below that a simple normalization by $(P_t(1-P_t))^{-1/2}$ makes this process stationary, thus enabling computation of N_{eV}^{eq} and F_{ST}^{eq}.

Interestingly, the decomposition (14) has some resemblance with cointegration of time series in econometrics (Granger, 1981, Engel and Granger, 1987). In our case there are s non-stationary time series P_{ti}, one for each subpopulation. If they are combined by means of a linear combination, the resulting cointegrated time series is stationary provided the vector of weights $\boldsymbol{v} = (v_1, \ldots, v_s)$ is orthogonal to $\mathbf{1}$ (see the appendix) and that the above mentioned normalization $(P_t(1 - P_t))^{-1/2}$ is applied.

Analogously to (14), we decompose the error term of (9) as $\boldsymbol{\varepsilon}_t = \varepsilon_t\mathbf{1} + \boldsymbol{\varepsilon}_t^0$, with $\varepsilon_t = \sum_{i=1}^{s} \gamma_i\varepsilon_{ti} = \boldsymbol{\gamma}\boldsymbol{\varepsilon}_t$. The recursion formula (9) is then reformulated as

$$\begin{aligned} P_{t+1} &= P_t + \varepsilon_{t+1}, \\ \boldsymbol{P}_{t+1}^0 &= \boldsymbol{B}\boldsymbol{P}_t^0 + \boldsymbol{\varepsilon}_{t+1}^0 = \boldsymbol{B}^0\boldsymbol{P}_t^0 + \boldsymbol{\varepsilon}_{t+1}^0, \end{aligned} \tag{16}$$

where in the second equation we utilized that $(\boldsymbol{B}\boldsymbol{P}_t)^0 = \boldsymbol{B}\boldsymbol{P}_t^0$ and introduced the matrix $\boldsymbol{B}^0$, that differs from $\boldsymbol{B}$ in that the eigenvalue of the right eigenvector $\mathbf{1}$ has been changed from 1 to 0, see the appendix for more details. The first part of (16) describes the dynamics of the genetic drift of the allele frequency of the whole population, whereas the second part describes the dynamics of the spatial subpopulation fluctuations of the allele frequencies.

For our purposes, it is enough to know the first and second moment properties of the quasi equilibrium distribution. To this end, in addition to (10) and (11), we assume that

$$E(\boldsymbol{P}_t^0|P_t) = \mathbf{0}, \tag{17}$$

corresponding to a selectively neutral allele. Moreover, in the appendix we motivate that

$$\boldsymbol{\Lambda}_t = \frac{\mathrm{Cov}(\boldsymbol{P}_t^0|P_t)}{P_t(1-P_t)} \to \boldsymbol{\Lambda} \tag{18}$$

and

$$\boldsymbol{\Sigma}_t = \frac{\mathrm{Cov}(\boldsymbol{\varepsilon}_{t+1}|P_t)}{P_t(1-P_t)} = \frac{E(\boldsymbol{\Sigma}(P_t\mathbf{1}+\boldsymbol{P}_t^0)|P_t)}{P_t(1-P_t)} \to \boldsymbol{\Sigma} \tag{19}$$

as t grows if the population is a large enough, so that P_t drifts slowly enough, using (10) in the second equality of (19).

Once $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ have been found, the quasi equilibrium values of the effective population size and fixation index satisfy the approximations

$$N_{eV}^{\mathrm{eq}} \approx \frac{1-(\boldsymbol{a}-\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{a}-\boldsymbol{\gamma})^T}{2\left(\boldsymbol{a}(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{\Lambda}(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{a}^T + \boldsymbol{a}\boldsymbol{\Sigma}\boldsymbol{a}^T\right)} \tag{20}$$

and

$$F_{ST}^{\mathrm{eq}} \approx \frac{\sum_{i=1}^s a_i\left((\boldsymbol{I}-\mathbf{1}\boldsymbol{a})\boldsymbol{\Lambda}(\boldsymbol{I}-\mathbf{1}\boldsymbol{a})^T\right)_{ii}}{1-(\boldsymbol{a}-\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{a}-\boldsymbol{\gamma})^T} \tag{21}$$

respectively, with $\boldsymbol{I}$ the identity matrix of order s, see the appendix for details.

5 Computational algorithm

Explicit expressions of N_{eV}^{eq} and F_{ST}^{eq} can be obtained for some very simple models. In general this is not possible though. Instead, we present an algorithm for evaluating (20) and (21) that is applicable for a wide range of models.

Let vech denote the half vectorization that transforms the lower part of a symmetric matrix of order s to a column vector of length $s(s+1)/2$. It is shown in the appendix that $\mathrm{vech}(\boldsymbol{\Sigma})$ and $\mathrm{vech}(\boldsymbol{\Lambda})$ satisfy the linear system of equations

$$\begin{array}{rcl}\mathrm{vech}(\boldsymbol{\Sigma}) & = & \boldsymbol{f} - \boldsymbol{F}\mathrm{vech}(\boldsymbol{\Lambda}), \\ \mathrm{vech}(\boldsymbol{\Lambda}) & = & \boldsymbol{G}\mathrm{vech}(\boldsymbol{\Sigma}),\end{array} \tag{22}$$

for some square matrices $\boldsymbol{F}$ and $\boldsymbol{G}$ of order $s(s+1)/2$ and column vector $\boldsymbol{f}$ of length $s(s+1)/2$. This gives an algorithm for computing N_{eV}^{eq} and F_{ST}^{eq} that involves the following steps:

1. Select total population size N and spatial structure, i.e. number of subpopulations s and migration matrix $\boldsymbol{M}$.
2. Compute $\boldsymbol{a}$ as the left eigenvector of $\boldsymbol{M}$ corresponding to the largest eigenvalue λ, normalized so that $\sum_{i=1}^{s} a_i = 1$, where $\lambda = 1$ because of (4). If not, replace $\boldsymbol{M}$ by $\boldsymbol{M}/\lambda$.
3. Compute the transition matrix $\boldsymbol{B}$ of the backward Markov chain from (5) and evaluate its equilibrium distribution $\boldsymbol{\gamma}$.
4. Compute $\boldsymbol{G}$ in (22) from (68)-(75) in the appendix.
5. Select reproduction scenario and compute $\boldsymbol{f}$ and $\boldsymbol{F}$ in (22) from (56) and (57) in the appendix. This is exemplified in (84), (86) and (88).
6. Find $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ by solving (22), i.e. evaluating $\mathrm{vech}(\boldsymbol{\Sigma}) = (\boldsymbol{J} + \boldsymbol{F}\boldsymbol{G})^{-1}\boldsymbol{f}$ and $\mathrm{vech}(\boldsymbol{\Lambda}) = \boldsymbol{G}\mathrm{vech}(\boldsymbol{\Sigma})$, where $\boldsymbol{J}$ is the identity matrix of order $s(s+1)/2$.
7. Evaluate N_{eV}^{eq} and F_{ST}^{eq} from (20) and (21).

The only input parameters of this algorithm are N, s and $\boldsymbol{M}$ of step 1 and the reproduction scenario (which may involve additional parameters) of step 5.

In the following two sections, we will illustrate the usefulness of (20) and (21) for several reproduction and migration models.

6 Reproduction scenarios

Reproduction scenario 1 (Exact migration proportions.) Suppose (3) holds exactly, so that $2Na_k m_{ki}$ of the genes of subpopulation k are transferred to subpopulation i in the next generation. Assume further that $\{\boldsymbol{\nu}_k^l = (\nu_{k1}^l, \ldots, \nu_{ks}^l)\}_{l=1}^{2Na_k}$ for each k is a collection of exchangeable random vectors.

We can then, without loss of generality, assume that genes $l = 1, \ldots, 2Na_kP_{tk}$ of subpopulation k have one of the alleles in generation t. The allele frequency of the genes that are passed on to subpopulation i is then

$$P^*_{tki} = \frac{1}{2Na_km_{ki}} \sum_{l=1}^{2Na_kP_{tk}} \nu^l_{ki}. \tag{23}$$

Combining the contribution to i from all subpopulations $k = 1, \ldots, s$, we thus have

$$P_{t+1,i} = \sum_{k=1}^{s} b_{ik} P^*_{tki}. \tag{24}$$

It turns out that the only further properties of ν^l_{ki} needed in order to compute $\mathbf{\Sigma}$ (see (86) in the appendix) are independence of $\{\boldsymbol{\nu}^l_k\}_{l=1}^{2Na_k}$ for different k and the covariances

$$V_{kij} = \mathrm{Cov}(\nu^l_{ki}, \nu^l_{kj}), \tag{25}$$

which do not depend on l because of the exchangeability assumption. □

Reproduction scenario 2 (Fertilization precedes migration.) Assume that an infinitely large gamete pool is constructed from gametes of individuals of subpopulation k and generation t, with allele frequency $\tilde{P}_{tk}$. We allow $\tilde{P}_{tk}$ to differ from P_{tk} to account for varying reproductive rate among the individuals of subpopulation k and model this by selecting $N_{ek} \leq Na_k$ breeders, which represents the local effective size of subpopulation k (see Hössjer et al., 2011). We assume that all $2N_{ek}$ breeding genes contribute in equal proportions to the gamete pool. If the genetic marker is selectively neutral, the allele frequency of the gamete pool has a hypergeometric distribution;

$$\tilde{P}_{tk}|P_{kt} \sim \mathrm{Hyp}(2Na_k, 2N_{ek}, P_{kt})/(2N_{ek}). \tag{26}$$

Then $2Na_km_{ki}$ genes are drawn independently (between genes and subpopulations i) from the gamete pool (26), with allele frequency

$$P^*_{tki}|\tilde{P}_{tk} \sim \mathrm{Bin}(2Na_km_{ki}, \tilde{P}_{tk})/(2Na_km_{ki}).$$

It is shown in the appendix that this is a special case of (23) with

$$V_{kij} \sim 1_{\{i=j\}}m_{ki} + m_{ki}m_{kj}\left(\frac{Na_k}{N_{ek}} - 1\right). \tag{27}$$

Notice that ν_{ki}^l and ν_{kj}^l, $i \neq j$, are uncorrelated (and in fact independent) when $N_{ek} = Na_k$, i.e. when all genes of subpopulation k become breeding genes. $\square$

Reproduction scenario 3 (Migration precedes fertilization.) For pollination of plants, it is of interest to reverse the order of migration and reproduction, so that gametes rather than individuals migrate. After migration, the new gamete pool of subpopulation i has allele frequency

$$\check{P}_{ti} = \sum_{k=1}^{s} B_{ik} \tilde{P}_{tk}, \tag{28}$$

with $\tilde{P}_{tk}$ as in (26). Then $2Na_i$ genes are drawn randomly from the post-migration gamete pool, yielding an allele frequency

$$P_{t+1,i} | \check{P}_{ti} \sim \text{Bin}(2Na_i, \check{P}_{ti})/(2Na_i) \tag{29}$$

of subpopulation i in the next generation $t+1$. When $B_{ik} = b_{ik}$ the gamete pools mix in exact proportions. More generally, we can allow for random mixing, where the rows

$$(B_{i1}, \ldots, B_{is}) \sim \text{Dir}\left(\alpha_i(b_{i1}, \ldots, b_{is})\right) \tag{30}$$

have independent Dirichlet distributions for $i = 1, \ldots, s$, so that $E(B_{ik}) = b_{ik}$ and

$$\text{Cov}(B_{ij}, B_{ik}) = \begin{cases} b_{ik}(1-b_{ik})/(\alpha_i+1), & j = k, \\ -b_{ij}b_{ik}/(\alpha_i+1), & j \neq k. \end{cases} \tag{31}$$

The parameter $\alpha_i \geq 0$ quantifies the amount of random variability of the mixing proportions when the contents of gamete pools from various subpopulations k migrate to subpopulation i. When $\alpha_i = 0$, all parents of subpopulation i are selected from the same randomly chosen subpopulation, with probabilities b_{ik}, whereas $\alpha_i = \infty$ gives exact mixing proportions $B_{ik} = b_{ik}$.

Notice that this model is not a special case of (25), since the remainder term of (3) is not zero, although it is asymptotically negligible when N and α_i are *both* large. However, exact mixing proportions alone ($\alpha_i = \infty$) are not sufficient for a vanishing remainder term in (3). $\square$

7 Spatial Structures

7.1 Equal reproductivity of all subpopulations.

The productivity of subpopulation k is defined as the average number of offspring

$$m_{k\cdot} = \sum_{i=1}^{s} m_{ki}$$

of its members. It follows from (4) that the average productivity of all populations, weighted according to their sizes, is

$$\sum_{k=1}^{s} a_k m_{k\cdot} = \sum_{k=1}^{s} a_k \sum_{i=1}^{s} m_{ki} = \sum_{i=1}^{s} a_i = 1, \tag{32}$$

a consequence of the constant total population size.

If all subpopulations are to be equally productive, it follows from (32) that $m_{k\cdot} = 1$ for all k. Then there is no systematic drift due to migration, i.e. $E(p_{t+1}|p_t) = p_t$, and $\boldsymbol{\gamma} = \boldsymbol{a}$, see for instance Hössjer (2011). Since $\boldsymbol{a}(\boldsymbol{B}-\boldsymbol{I}) = \boldsymbol{\gamma}(\boldsymbol{B}-\boldsymbol{I}) = \mathbf{0}$, it follows that (20) simplifies to

$$N_{eV}^{\text{eq}} = \frac{1}{2\boldsymbol{a}\boldsymbol{\Sigma}\boldsymbol{a}^T}. \tag{33}$$

Moreover, the definition of $\boldsymbol{\Lambda}$ implies $\boldsymbol{a}\boldsymbol{\Lambda} = \boldsymbol{\gamma}\boldsymbol{\Lambda} = \mathbf{0}$, and therefore (21) simplifies to

$$F_{ST}^{\text{eq}} = \sum_{i=1}^{s} a_i \Lambda_{ii}. \tag{34}$$

For reproduction scenario 2, if all local effective sizes equal the local census sizes ($N_{ek} = Na_k$), it follows from calculations in the appendix (cf. (84)) that $\boldsymbol{\Sigma}$ is a diagonal matrix with entries

$$\Sigma_{ii} = \frac{1}{2Na_i}\left(1 - \sum_{k=1}^{s} b_{ik}\Lambda_{kk}\right) \tag{35}$$

Inserting this expression into (33) and using $\sum_{i=1}^{s} a_i b_{ik} = a_k$, since $\boldsymbol{a} = \boldsymbol{\gamma}$, we obtain

$$N_{eV}^{\text{eq}} = \frac{1}{2\sum_{i=1} a_i^2 \Sigma_{ii}} = \frac{N}{1 - \sum_{k=1}^{s} a_k \Lambda_{kk}} = \frac{N}{1 - F_{ST}^{\text{eq}}}, \tag{36}$$

the well known formula originally derived by Wright (1951) for the island model. Wang and Caballero (1999, eqn. 15) showed the validity of (36) more generally when fertilization precedes migration.

For reproduction scenario 3, a similar calculation when $N_{ek} = Na_k$ (see (88)) yields

$$N_{eV}^{\text{eq}} = \frac{N}{1 - \sum_{i=1}^{s} a_i (\boldsymbol{B\Lambda B})_{ii} + \sum_{k=1}^{s} \Lambda_{kk} \sum_{i=1}^{s} \frac{a_i^2}{1+\alpha_i} b_{ik} - \sum_{i=1}^{s} \frac{a_i^2}{1+\alpha_i} (\boldsymbol{B\Lambda B})_{ii}}. \tag{37}$$

However, neither (36) nor (37) hold when $N_{ek} \neq Na_k$, since $\boldsymbol{\Sigma}$ is then no longer diagonal.

Spatial structure 1 (Island model.) The most well known population genetic model with spatial structure is the island model (Wright, 1943, Maruyama, 1970b). All subpopulations (or islands) have equal size, $a_k = 1/s$, and migration is symmetric, so that a fraction m of the offspring select island uniformly (including its present island), and the remaining fraction $1-m$ of offspring never migrate. This corresponds to

$$\boldsymbol{M} = \boldsymbol{B} = (1-m)\boldsymbol{I} + \frac{m}{s}\boldsymbol{1}\boldsymbol{1}^T, \tag{38}$$

an overall migration rate $m' = m(s-1)/s$ (cf. (6)) and an average relative size $|\mathcal{N}| = 1$ of the one-step neighborhood in (7). In general it is not possible to find explicit expressions for N_{eV}^{eq} and F_{ST}^{eq}. However, due its symmetry, the island model is an exception.

Consider reproduction scenario 2, with a constant local effective population size $N_{ek} = N_e$ for all islands. Then, it is shown in the appendix that (20) and (21) simplify to

$$N_{eV}^{\text{eq}} = \frac{sN_e}{1 - F_{ST}^{\text{eq}}} \tag{39}$$

and

$$F_{ST}^{\text{eq}} = \frac{1}{\frac{s}{s-1} 2\tilde{N}\left(1 - (1-m)^2\right) + 1} \tag{40}$$

respectively, with

$$\frac{1}{\tilde{N}} = \frac{(1-m)^2}{N_e} + \frac{2m - m^2}{N/s} \tag{41}$$

a weighted harmonic average of N_e and N/s. Formula (40) was derived in Hössjer et al. (2011) by other methods, exploiting that $F_{ST,t}$ is a univariate autoregressive time series for the island model. Related expressions for F_{ST}^{eq} can also be found in Wright (1943), Nei (1975), Takahata (1983), Takahata and Nei (1984) and Ryman and Leimar (2008).

For reproduction scenario 3, with $N_{ek} = N_e$ and $\alpha_i = \alpha$, the corresponding formulas

$$N_{eV}^{\text{eq}} = \frac{sN_e}{1 - F_{ST}^{\text{eq}} + \left(\frac{N_e}{N/s} + \frac{2N_e}{\alpha+1}\right)\left(1-(1-m)^2\right)F_{ST}^{\text{eq}}} \tag{42}$$

and

$$F_{ST}^{\text{eq}} = \frac{1}{(\frac{s}{s-1} - \frac{1}{\alpha+1})2\tilde{N}\left(1-(1-m)^2\right) + \frac{\tilde{N}}{N_e}(1-m)^2} \tag{43}$$

are valid when α is large or $N_e = N/s$. Formula (42) is, to the best of our knowledge, new, and (43) extends a formula of Hössjer et al. (2011) to $\alpha \neq 0$.

□

Spatial structure 2 (Stepping stone models.) In natural populations, migration is often restricted to neighbouring subpopulations. Kimura (1953) proposed a class of stepping stone models that reflect this behaviour. Its first mathematical treatment appeared in Kimura and Weiss (1964) and Weiss and Kimura (1965). When $s \geq 3$, the one-dimensional circular version of the stepping stone model has all subpopulations distributed on a circle with

$$m_{ki} = \begin{cases} 1-m, & i = k, \\ m/2, & i = k+1 \text{ or } k-1 \text{ modulo } s, \\ 0, & \text{otherwise.} \end{cases} \tag{44}$$

When $s = 2$, the nonzero offdiagonal elements are instead $m_{12} = m_{21} = m$, since $k+1 = k-1$ modulo 2. Maruyama (1970a) considered the eigenvalue effective population size N_{eE}^{eq} when either $m \to 0$ or $N \to \infty$. Based on his asymptotic results, Wang and Caballero (1999) suggested the approximation

$$N_{eE}^{\text{eq}} \approx sN_e + \frac{s^2}{2m\pi^2}, \tag{45}$$

for all values of m, when $N_{ek} = N_e = N/s$ and s is an even integer.

A linear stepping stone model can be defined when $s \geq 3$ that differs from (44) in that no direct communication occurs between the end populations 1 and s, i.e. $m_{1s} = m_{s1} = 0$. All other off-diagonal elements of $\boldsymbol{M}$ are given by (44), whereas the diagonal elements satisfy $m_{11} = m_{ss} = 1 - m/2$ and $m_{ii} = 1 - m$ for $2 \leq i \leq s-1$.

A two-dimensional stepping stone model has $s = s_1 s_2$ subpopulations $\{i = (i_1, i_2);\ 1 \leq i_1 \leq s_1, 1 \leq i_2 \leq s_2\}$ positioned on a rectangular grid, with

$s_1 \geq 2$ and $s_2 \geq 2$. In order to avoid edge effects we identify islands along the left edge as neighbours to those along the right edge, provided their second coordinates agree. Similarly, we identify islands along the bottom and top edges as neighbours whenever their first coordinates agree. Consequently, the islands can be thought of as uniformly positioned along a torus with migration intensities

$$m_{(k_1,k_2),(i_1,i_2)} = \begin{cases} 1-m, & (i_1,i_2) = (k_1,k_2), \\ m/4, & i_2 = k_2 \text{ and } i_1 = k_1+1 \text{ or } k_1-1 \text{ modulo } s_1, \\ m/4, & i_1 = k_1 \text{ and } i_2 = k_2+1 \text{ or } k_2-1 \text{ modulo } s_2, \\ 0, & \text{otherwise}, \end{cases} \tag{46}$$

when $s_1 \geq 3$ and $s_2 \geq 3$. When s_1 and/or s_2 equals 2, the off-diagonal entries $m_{(k_1,k_2),(i_1,i_2)}$ are added for those (i_1,i_2) that correspond to the same subpopulation.

When $s_1 \geq 3$ and $s_2 \geq 3$, a rectangular version without edge effects is defined by putting those off-diagonal elements in $\boldsymbol{M}$ that correspond to transitions between opposite edges. Hence, the diagonal elements are $1-m$ for interior subpopulations, $1-3m/4$ along edges and $1-m/2$ at corners.

It is easily verified that $\boldsymbol{B} = \boldsymbol{M}$, $\boldsymbol{a} = \boldsymbol{\gamma} = \boldsymbol{1}^T/s$ for all four stepping stone models. After some computation, this implies that the average relative size of the one step neighborhood in increasing order is $|\mathcal{N}| = \min(1, 2/(s-1))$ for the linear model, $|\mathcal{N}| = 2/s$ for the circular model, $|\mathcal{N}| = (4-2(s_1^{-1} + s_2^{-1}))/(s-1)$ for the rectangular model and $|\mathcal{N}| = (4-1_{\{|s_1|=2\}}-1_{\{|s_2|=2\}})/(s-1)$ for the torus model. The overall migration rate is $m' = m$ for the circular and torus models, $m' = m(s-1)/s$ for the linear model and $m(1-0.5(s_1^{-1} + s_2^{-1}))$ for the rectangular model. See Durrett (2008) for more results on stepping stone models. □

7.2 Varying reproductivity of subpopulations

In general the reproductivity $m_{k\cdot}$ varies between members of different subpopulations k. The migration matrix may then be expressed as

$$\boldsymbol{M} = \boldsymbol{R}\boldsymbol{F}, \tag{47}$$

with $\boldsymbol{R} = \operatorname{diag}(m_{1\cdot}, \ldots, m_{s\cdot})$ a diagonal matrix of reproductivities, $\boldsymbol{F} = (f_{ki})_{k,i=1}^{s}$ the transition matrix of a forward Markov chain and f_{ki} the probability that a copy of a subpopulation k gene that is passed on to the next generation ends up in subpopulation i.

Spatial structure 3 (Genetic reservoir.) Suppose $s \geq 2$ and that one subpopulation s acts a genetic reservoir for the other subpopulations $1, \ldots, s-1$. The reservoir is located in the center of a circle and the other subpopulations symmetrically along its perimeter. We assume that the productivities $m_{1\cdot} = \ldots = m_{s-1,\cdot}$ and sizes $a_1 = \ldots = a_{s-1}$ of the perimeter populations are identical, with

$$0 < m_{1\cdot} < 1 < m_{s\cdot} \tag{48}$$

determining how much more productive the reservoir is compared to the other subpopulations. It follows from (1) and (32) that

$$a_1 = \frac{1}{s-1}\frac{m_{s\cdot}-1}{m_{s\cdot}-m_{1\cdot}} \text{ and } a_s = \frac{1-m_{1\cdot}}{m_{s\cdot}-m_{1\cdot}}.$$

When $s \geq 4$, the entries of the forward transition matrix are specified as

$$f_{ki} = \begin{cases} \beta, & k = s, i = 1, \ldots, s-1, \\ 1-(s-1)\beta, & k = i = s, \\ \gamma, & k = 1, \ldots, s-1, i = s, \\ \delta/2, & k = 1, \ldots, s-1, i = k-1 \text{ or } k+1 \text{ modulo } s-1, \\ 1-\gamma-\delta, & k = i = 1, \ldots, s-1, \\ 0, & \text{otherwise.} \end{cases}$$

When $s = 3$, the migration rates for $i = k-1$ and $i = k+1$ modulo $s-1$ are added together, since they correspond to the same subpopulation i. When $s = 2$, we put $\delta = 0$. The perimeter islands act as a circular stepping stone model with migration from the reservoir and internal probability δ that any gene migrates to a neighbouring perimeter island. The other two parameters γ and β control the degree of gene flow between the reservoir and the perimeter populations. It turns out that in order for (4) to hold, β and γ must be related as

$$(s-1)a_s m_{s\cdot}\beta = a_s(m_{s\cdot}-1) + (s-1)a_1 m_{1\cdot}\gamma,$$

where the left hand side is the total migration rate from the reservoir to the perimeter, the first term on the right hand side is the migration rate from the reservoir needed to keep its population size constant in absence of immigration, and the last term is the total migration rate from the perimeter to the reservoir.

Hence there are four parameters; $m_{1\cdot}$, $m_{s\cdot}$, δ and γ that can be varied in this model subject to constraints (48), $0 \leq \delta \leq 1$ and $0 < \gamma \leq \min(1-\delta, (1-$

$m_{1\cdot})/(m_{1\cdot}(m_{s\cdot}-1)))$. Given δ, the upper bound on γ gives the maximal possible gene flow between the reservoir and perimeter, and it assures that all diagonal elements of $\boldsymbol{F}$ are non-negative. The total migration rate (6) is

$$m' = (s-1)\left(a_1 m_{1\cdot}(\gamma+\delta) + a_s m_{s\cdot}\beta\right).$$

Because of (48), we typically have $E(p_{t+1}|p_t) \neq p_t$ and $\boldsymbol{\gamma} \neq \boldsymbol{a}$. $\square$

8 Multilocus and multiallelic extension

Assume there are L genetic markers, with the l:th marker having alleles $q = 1, \ldots, n_l$ for $l = 1, \ldots, L$. Let P_{ti}^{lq} and $p_t^{lq} = \sum_{i=1}^{s} a_i P_{ti}^{lq}$ refer to the frequency of allele q at marker l in subpopulation i and the whole population respectively. If there was no differentiation among subpopulations, the fraction of heterozygots in generation t would be

$$H_{Tt} = \frac{1}{L}\sum_{l=1}^{L}\left(1 - \sum_{q=1}^{n_l}(p_t^{lq})^2\right),$$

whereas the expected fraction of heterozygots, averaged over all subpopulations, in generation t, is

$$H_{St} = \sum_{i=1}^{s} a_i \frac{1}{L}\sum_{l=1}^{L}\left(1 - \sum_{q=1}^{n_l}(P_{ti}^{lq})^2\right).$$

The coefficient of gene differentiation

$$G_{ST,t} = \frac{H_{Tt} - H_{St}}{H_{Tt}} = \frac{\sum_{lq}\sum_{i=1}^{s} a_i (P_{ti}^{lq} - p_t^{lq})^2}{\sum_{lq} p_t^{lq}(1 - p_t^{lq})} \tag{49}$$

is a multiallelic and multilocus extension of $F_{ST,t}$. In the appendix, it is shown that (21) is a valid quasi equilibrium formula for $G_{ST,t}$ as well, i.e.

$$G_{ST}^{\text{eq}} \approx \frac{\sum_{i=1}^{s} a_i \left((\boldsymbol{I} - \mathbf{1}\boldsymbol{a})\boldsymbol{\Lambda}(\boldsymbol{I} - \mathbf{1}\boldsymbol{a})^T\right)_{ii}}{1 - (\boldsymbol{a} - \boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{a} - \boldsymbol{\gamma})^T}. \tag{50}$$

Indeed, we argue that the the approximation (50) is increasingly accurate the larger the number of loci L is.

9 Numerical Results

In Figure 1 we display N_{eV}^{eq} and F_{ST}^{eq} as functions of the migration rate m' in (6) for the island and various stepping stone models when fertilization precedes migration. In general, the larger the average relative size $|\mathcal{N}|$ of the one step neighborhood (cf. (7)) is, the smaller N_{eV}^{eq} and F_{ST}^{eq} are. However, N_{eV}^{eq} varies very little between the linear stepping stone model (with smallest $|\mathcal{N}|$) and the island model (with largest $|\mathcal{N}|$). The fixation index F_{ST}^{eq} varies somewhat more, at least in relative terms. We also notice that although N_{eV}^{eq} and F_{ST}^{eq} vary not only with the local effective size N_e, but also with the local census size Na_k, the dependence on the latter is very small (see also Hössjer et al., 2011).

When migration precedes fertilization, the impact of the Dirichlet parameter $\alpha_i = \alpha$ in (30) is quite dramatical, as shown in Figure 2. Even though subpopulation sizes are kept fixed, a larger variability of migration proportions (a smaller α) implies a substantially decreased N_{eV}^{eq} as well as a lowered F_{ST}^{eq}. As in Figure 1 we conclude that the migration structure (island or stepping stone) has less effect.

In Figure 3 we compare N_{eV}^{eq} with the eigenvalue effective population size N_{eE}^{eq} in (45) for the circular stepping stone model. The agreement is quite good, although for small migration rates N_{eE}^{eq} is somewhat larger.

Quasi equilibrium is approached at a rate determined by the eigenvalue of the matrix $\boldsymbol{B}^0$ in (16) with largest modulus. In the appendix, it is shown that this so called spectral radius of $\boldsymbol{B}^0$ equals $|\lambda_2| < 1$, the second largest modulus of the eigenvalue of $\boldsymbol{B}$. It is close to 1 for moderate amounts of migration and $-\log(|\lambda_2|) \approx 1 - |\lambda_2|$ is the rate at which quasi equilibrium is attained, which takes time $O\left((1-|\lambda_2|)^{-1}\right)$. In Figure 4 it is shown that $1 - |\lambda_2|$ is proportional to the migration rate m' for the island and several stepping stone models. However, the proportionality constant varies quite a lot between models. The larger $|\mathcal{N}|$ is, the faster is the convergence to quasi equilibrium.

Figure 5 shows, for the genetic reservoir model, that N_{eV}^{eq} and F_{ST}^{eq} are both very sensitive to varying reproductivity between subpopulations. The larger the migration rate $m_{s\cdot}$ of the reservoir, the smaller are N_{eV}^{eq} and F_{ST}^{eq}. For instance, when $m_{s\cdot} = 20$, the reservoir occupies $a_s = 4.76\%$ of the total population. Still, almost all genes are inherited from the reservoir, so that N_{eV}^{eq} approximately equals the number of individuals Na_s in it. On the other hand, the migration rate δ between the perimeter islands and the migration rate γ from the perimeter to the reservoir effect N_{eV}^{eq} and F_{ST}^{eq} very little.

10 Discussion

In this paper, we have developed a general methodology for joint computation of the fixation index and variance effective size of a large class of populations exhibiting substructure. Conditioning on that no allele gets fixed in the population, we decompose the vector valued time series of subpopulation allele frequencies into two parts, corresponding to genetic drift of the whole population and genetic differentiation between subpopulations respectively. The latter is stationary modulo a simple standardization, and it enables the quasi equilibrium values of the fixation index and variance effective population size to be computed as functions of the standardized covariance matrices of the genetic drift ($\mathbf{\Sigma}$) and the spatial allele frequency fluctuations ($\mathbf{\Lambda}$).

The numerical illustrations reveal that N_{eV}^{eq} and F_{ST}^{eq} are both very sensitive to systematic variation of reproductivities between subpopulations, or random variation of migration proportions between subpopulations. Apart from that, a few parameters (such as s, m' and N_e) seem to characterize N_{eV}^{eq} and F_{ST}^{eq} quite well, whereas the exact spatial pattern and neighborhood structure is less important.

Our approach relies on two approximations. First, once a mutation has occurred at a given locus, we ignore the possibility of new mutations before quasi equilibrium is attained. Thus the mutation rate has to be smaller than the convergence rate $1 - |\lambda_2|$ to quasi equilibrium, where $|\lambda_2|$ is the second largest modulus of the possibly complex eigenvalues of $\boldsymbol{B}$. For the island model, this entails a mutation rate smaller than $1 - |\lambda_2| = m = sm'/(s-1)$. For other models, with smaller average relative size $|\mathcal{N}|$ of the one step neighborhoods, $1-|\lambda_2|$ is typically smaller but still of the same order of magnitude as the migration rate m'. In any case, since mutation rates are thought to be of the order of $10^{-4} - 10^{-3}$ for highly polymorphic markers such as microsatellites and $10^{-7} - 10^{-6}$ for allozymes and single nucleotide polymorphisms (Waples and Gaggiotti 2006; Allendorf and Luikart 2007), this puts very mild restrictions on the migration rates. Secondly, we have simplified the algorithm for computing $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$ by means of the large population approximation

$$N_{eV}(1 - |\lambda_2|) \gg 1, \tag{51}$$

which implies that the amount of genetic drift or loss of heterozygosity is small during the time it takes for quasi equilibrium to be attained. This seems reasonable for most migration models unless the migration rate and number of subpopulations are both very small. In principle, more complicated expressions can be worked out for $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$ by dropping (51). For the island model, this has been done by Hössjer et al. (2012).

A number of extensions are possible. Firstly, we could apply our approach to populations that simultaneously accommodate spatial substructure and overlapping generations. If the population consists of $s = s_1 s_2$ subpopulations $\{i = (i_1, i_2);\ 1 \leq i_1 \leq s_1, 1 \leq i_2 \leq s_2\}$, with geographic sites i_1 and age groups i_2,

$$F_{ST,t} = \frac{\sum_{i_1=1}^{s_1} a_{i_1\cdot}(P_{ti_1\cdot} - p_t)^2}{p_t(1-p_t)}$$

is a more natural definition of the fixation index than (13), with $a_{i_1\cdot} = \sum_{i_2=1}^{s_2} a_{(i_1,i_2)}$ and $P_{ti_1\cdot} = \sum_{i_2=1}^{s_2} P_{t,(i_1,i_2)}$. The fixation index F_{ST}^{eq} is still a function of the standardized covariance matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$, although a different one than (21). The variance effective size N_{eV}^{eq} is either computed from (20), as the loss of genetic heterogeneity *per time unit*. Alternatively, to obtain the genetic loss *per generation*, one has to multiply the denominator of (20) by the mean generation time, see Engen et al. (2005a).

Secondly, (2) could be dropped, allowing the local as well as the total population sizes to fluctuate in time. For instance, Whitlock and Baron (1997) and Nunney (1999) derive inbreeding effective population sizes under such assumptions. In our setting of a variance effective population size, one could allow $\boldsymbol{N}_t = (N_{t1}, \ldots, N_{ts})$, the vector of subpopulation sizes at time t, to satisfy a recursion

$$\boldsymbol{N}_{t+1} = \boldsymbol{N}_t \boldsymbol{M}_t$$

where $\boldsymbol{M}_t$ is a stochastic Leslie matrix with $E(\boldsymbol{M}_t) = \boldsymbol{M}$ involving demographic, genetic as well as environmental effects, see Engen et al. (2005a,b). The corresponding recursion

$$\begin{aligned} \boldsymbol{P}_{t+1} &= \boldsymbol{B}_t \boldsymbol{P}_t + \boldsymbol{\varepsilon}_{t+1} \\ &= \boldsymbol{B} \boldsymbol{P}_t + \boldsymbol{\epsilon}_{t+1}, \end{aligned} \tag{52}$$

of the subpopulation allele frequencies is slightly more complicated than (9). The first error term $\boldsymbol{\varepsilon}_t$ involves the reproduction scheme, and is assumed to satisfy $E(\boldsymbol{\varepsilon}_t) = 0$ for a selectively neutral allele. The second error term $\boldsymbol{\epsilon}_{t+1} = (\boldsymbol{B}_t - \boldsymbol{B})\boldsymbol{P}_t + \boldsymbol{\varepsilon}_{t+1}$ involves both reproduction and stochastic migration rates. The entries of the backward Markov transition matrix $\boldsymbol{B}_t$ are defined similarly to (5), but as functions of $\boldsymbol{M}_t$ and $\boldsymbol{N}_t$ rather than $\boldsymbol{M}$ and $\boldsymbol{a}$. Under small perturbations $\boldsymbol{M}_t - \boldsymbol{M}$, we should have $E(\boldsymbol{B}_t) \approx \boldsymbol{B}$ and hence $E(\boldsymbol{\epsilon}_{t+1}) \approx 0$. If so, one can proceed as before, computing jointly the covariance matrices of $\boldsymbol{\epsilon}_{t+1}$ and $\boldsymbol{P}_t^0$ conditional on P_t, and, finally, expressions for F_{ST}^{eq} and N_{eV}^{eq}.

Thirdly, the numerical algorithm in Section 5 involves solving the linear system (22) of equations with $s(s+1)$ unknown variables. For this reason, it is computationally feasible only for small s. Although the number of unknown variables can be reduced to $s(s+1)/2+s$ when $\boldsymbol{\Sigma}$ is diagonal (as in (35)), the complexity is still of the same order of magnitude. Therefore, faster algorithms would be of great use. If the entries of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ only depend on the distance between subpopulations (as for the circular and torus stepping stone models), the number of parameters in (22) is reduced to s.

Fourthly, the spatial covariance matrix $\boldsymbol{\Lambda}$ is interesting in its own right, since it enables quasi equilibrium autocorrelations

$$\rho_{ij} = \frac{\Lambda_{ij}}{\sqrt{\Lambda_{ii}\Lambda_{jj}}} \tag{53}$$

to be computed analytically between all pairs of subpopulations i and j. It would be of great interest to compare (53) with simulation based tools for spatial correlation, see Hardy and Vekemans (2002).

Appendix.

Jordan decomposition of $\boldsymbol{B}$ and motivation of (16). Let $\boldsymbol{B} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1}$ be the Jordan canonical form of $\boldsymbol{B}$, with

$$\boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_1 & \cdots & & \\ 0 & \ddots & & \vdots \\ \vdots & & & \vdots \\ 0 & \cdots & 0 & \boldsymbol{D}_r \end{pmatrix}$$

a block diagonal matrix containing the (possibly complex-valued) eigenvalues of $\boldsymbol{B}$ along the diagonal. For each $l = 1,\ldots,r$, the square matrix $\boldsymbol{D}_l$ occupies rows and columns $j_{l-1}+1,\ldots,j_l$ of $\boldsymbol{D}$, with diagonal entries equal to λ_l, all entries along the superdiagonal equal to 1 and all other entries of $\boldsymbol{D}_l$ equal 0. Hence λ_l is an eigenvalue of $\boldsymbol{B}$ of multiplicity $j_l - j_{l-1}$, with $0 = j_0 < j_1 < \ldots < j_r = s$. In particular, $\boldsymbol{D}$ is diagonal when all egienvalues of $\boldsymbol{B}$ are distinct and $r = s$. Then the rows of $\boldsymbol{V}^{-1}$ contain the left eigenvectors of $\boldsymbol{B}$ and the columns $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_s$ of $\boldsymbol{V}$ the right eigenvectors. See for instance Cox and Miller (1965).

In any case, regardless of whether $\boldsymbol{D}$ is diagonal or not, since $\boldsymbol{B}$ is a transition matrix of a Markov chain, $\boldsymbol{v}_1 = \mathbf{1}$ is a right eigenvector with eigenvalue $\lambda_1 = 1$. By the assumed irreducibility and aperiodicity of this Markov chain, it follows from the Perron Frobenius Theorem that $|\lambda_j| < 1$ for $j = 2,\ldots,r$, and without loss of generality, we may assume $|\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_r| \geq 0$.

Introduce the scalar product $(\boldsymbol{u},\boldsymbol{v})=\sum_{i=1}^{s}\gamma_i\bar{u}_iv_i$ for possibly complex-valued column vectors $\boldsymbol{u}=(u_i)$ and $\boldsymbol{v}=(v_i)$ of length s, with $\bar{u}_i$ the complex conjugate of u_i. It can be shown that $\boldsymbol{v}_2,\ldots,\boldsymbol{v}_s$ are all orthogonal to $\boldsymbol{v}_1$ with respect to $(\cdot,\cdot)$. Indeed, if $j=j_{l-1}+1$ and $l=2,\ldots,r$, we have $\boldsymbol{B}\boldsymbol{v}_j=\lambda_l\boldsymbol{v}_j$, so that

$$\begin{array}{rcl}\lambda_l(\boldsymbol{v}_1,\boldsymbol{v}_j) & = & (\boldsymbol{v}_1,\boldsymbol{B}\boldsymbol{v}_j)\\ & = & \sum_{i=1}^{s}\gamma_i\bar{\boldsymbol{v}}_{1i}(\boldsymbol{B}\boldsymbol{v}_j)_i\\ & = & \sum_{i=1}^{s}\gamma_i\sum_{k=1}^{s}b_{ik}v_{jk}\\ & = & \sum_{k=1}^{s}v_{jk}\sum_{i=1}^{s}\gamma_ib_{ik}\\ & = & \sum_{k=1}^{s}\gamma_kv_{jk}\\ & = & (\boldsymbol{v}_1,\boldsymbol{v}_j),\end{array}$$

and hence $(\boldsymbol{v}_1,\boldsymbol{v}_j)=0$ since $\lambda_l\neq 1$. In the second last step we used that $\boldsymbol{\gamma}$ is the asymptotic distribution of $\boldsymbol{B}$. If $j_{l-1}+2\leq j\leq j_l$, we have $\boldsymbol{B}\boldsymbol{v}_j=\lambda_l\boldsymbol{v}_j+\boldsymbol{v}_{j-1}$, and a similar calculation yields

$$(\boldsymbol{v}_1,\boldsymbol{v}_j)=(\boldsymbol{v}_1,\boldsymbol{B}\boldsymbol{v}_j)=(\boldsymbol{v}_1,\boldsymbol{v}_{j-1})+\lambda_l(\boldsymbol{v}_1,\boldsymbol{v}_j)=\lambda_l(\boldsymbol{v}_1,\boldsymbol{v}_j),$$

using induction with respect to j in the last step. Since $\lambda_l\neq 1$, we conclude again that $(\boldsymbol{v}_1,\boldsymbol{v}_j)=0$.

It follows that

$$\boldsymbol{B}\boldsymbol{v}=\boldsymbol{B}(v\boldsymbol{1}+\boldsymbol{v}^0)=v\boldsymbol{1}+\boldsymbol{B}\boldsymbol{v}^0,$$

for any vector $\boldsymbol{v}$, with $v=(\boldsymbol{v},\boldsymbol{1})$. Since $\boldsymbol{v}^0$ is a linear combination of $\boldsymbol{v}_2,\ldots,\boldsymbol{v}_s$, so is $\boldsymbol{B}\boldsymbol{v}^0$, and hence orthogonal to $\boldsymbol{1}$. Consequently, $\boldsymbol{B}\boldsymbol{v}^0=(\boldsymbol{B}\boldsymbol{v})^0$.

Define $\boldsymbol{D}_0=\operatorname{diag}(\lambda_1^0,\lambda_2,\ldots,\lambda_s)$ by replacing the eigenvalue $\lambda_1=1$ in $\boldsymbol{D}$ by $\lambda_1^0=0$ (or any other value $-|\lambda_2|\leq\lambda_1^0\leq|\lambda_2|$), and put

$$\boldsymbol{B}^0=\boldsymbol{V}\boldsymbol{D}^0\boldsymbol{V}^{-1}. \tag{54}$$

Denote by $\boldsymbol{e}_i=(0,\ldots,0,1,0,\ldots,0)$ the unit vector with 1 in position i and zeros elsewhere. Since $\boldsymbol{V}^{-1}\boldsymbol{v}^0$ is linear combination of $\boldsymbol{e}_2^T,\ldots,\boldsymbol{e}_s^T$, it follows that $\boldsymbol{D}\boldsymbol{V}^{-1}\boldsymbol{v}^0=\boldsymbol{D}^0\boldsymbol{V}^{-1}\boldsymbol{v}^0$, and hence $\boldsymbol{B}\boldsymbol{v}^0=\boldsymbol{B}^0\boldsymbol{v}$. $\square$

Motivating (18), (19) and (22). We will assume that the covariance matrix (11) of the genetic drift is a quadratic form

$$\Sigma_{ij}(\boldsymbol{P}_t)=f_{ij}P_t(1-P_t)-\sum_{1\leq u\leq v}F_{ij,uv}P_{tu}^0P_{tv}^0+\sum_{u=1}^{s}D_{ij,u}P_{tu}^0 \tag{55}$$

in $\boldsymbol{P}_t$, where $\boldsymbol{P}_t^0=(P_{t1}^0,\ldots,P_{ts}^0)$. Indeed, (55) is satisfied for all reproduction scenarios studied in this paper (see below). Conditioning on P_t and taking expectation with respect to $\boldsymbol{P}_t^0$, (55) and (19) imply

$$\Sigma_{t,ij} = f_{ij} - \sum_{1\le v\le u} F_{ij,uv}\Lambda_{t,uv}. \tag{56}$$

or, using vector notation,

$$\text{vech}(\boldsymbol{\Sigma}_t) = \boldsymbol{f} - \boldsymbol{F}\text{vech}(\boldsymbol{\Lambda}_t), \tag{57}$$

where $\Sigma_{t,ij}$ refers to component number $(j-1)s+i-j(j-1)/2$ of vech$(\boldsymbol{\Sigma}_t)$, assuming $i \ge j$ and that the elements of $\boldsymbol{\Sigma}_t$ are listed columnwise from column 1 (with s entries) to column s (with one entry). The elements of the column vectors $\boldsymbol{f} = (f_{ij})$ and vech$(\boldsymbol{\Lambda}_t)$ are listed in the same way, and $\boldsymbol{F} = (F_{ij,uv})$ a square matrix of order $s(s+1)/2$, cf. the upper part of (22).

Next we derive a recursion formula for $\boldsymbol{\Lambda}_t$. As a preliminary, we start by introducing the effective population size

$$N_{eV,t} = \frac{P_t(1-P_t)}{2\text{Var}(P_{t+1}-P_t|P_t)} = \frac{1}{2\boldsymbol{\gamma}\boldsymbol{\Sigma}_t\boldsymbol{\gamma}^T}.$$

that is based on the drift of P_t rather than p_t, as in (12). It can be motivated that the drift backwards in time satisfies

$$E(P_{t-1}-P_t|P_t) = O\left(N_{eV,t-1}^{-1}\right) \tag{58}$$

and

$$E\left((P_{t-1}-P_t)^2|P_t\right) = \frac{P_t(1-P_t)}{2N_{eV,t-1}}\left(1+O(N_{eV,t-1}^{-1})\right). \tag{59}$$

Indeed, by the central limit theorem, we can justify that $\varepsilon_t|P_{t-1}$ has an approximately normal distribution $N(0,\sigma^2P_{t-1}(1-P_{t-1}))$ for small $\sigma^2 = (2N_{eV,t-1})^{-1}$. Denoting the prior density of P_{t-1} by $f_{P_{t-1}}$, Bayes' formula implies

$$f_{(P_{t-1}-P_t)/\sigma|P_t}(y) = \frac{g(P_t+\sigma y)}{C(\sigma)}\exp\left(-\frac{y^2}{2(P_t+\sigma y)(1-P_t-\sigma y)}\right),$$

where $g(x) = f_{P_{t-1}}(x)/(\sqrt{2\pi x(1-x)})$ and $C(\sigma)$ is a normalizing constant assuring that $f_{(P_{t-1}-P_t)/\sigma|P_t}$ integrates to 1. By a Taylor series expansion with respect to σ, it then follows that

$$\begin{array}{rcl} C(\sigma) & = & C(0)+O(\sigma^2), \\ \int y f_{(P_{t-1}-P_t)/\sigma|P_t}(y)dy & = & O(\sigma), \\ \int y^2 f_{(P_{t-1}-P_t)/\sigma|P_t}(y)dy & = & P_t(1-P_t)(1+O(\sigma^2)), \end{array}$$

thus motivating (58) and (59).

We then plug the lower part of (16) into (18) and obtain

$$\begin{array}{rcl} \boldsymbol{\Lambda}_t & = & \mathrm{Cov}(\boldsymbol{B}^0\boldsymbol{P}^0_{t-1} + \boldsymbol{\varepsilon}^0_t | P_t)/(P_t(1-P_t)) \\ & = & \left(\boldsymbol{B}^0\mathrm{Cov}(\boldsymbol{P}^0_{t-1}|P_t)(\boldsymbol{B}^0)^T + \mathrm{Cov}(\boldsymbol{\varepsilon}^0_t|P_t)\right)/(P_t(1-P_t)), \end{array} \tag{60}$$

where

$$\begin{array}{rcl} \mathrm{Cov}(\boldsymbol{P}^0_{t-1}|P_t) & = & E\left(\mathrm{Cov}(\boldsymbol{P}^0_{t-1}|P_{t-1},P_t)|P_t\right) + \mathrm{Cov}\left(E(\boldsymbol{P}^0_{t-1}|P_{t-1},P_t)|P_t\right) \\ & = & E\left(\mathrm{Cov}(\boldsymbol{P}^0_{t-1}|P_{t-1})|P_t\right) + \mathrm{Cov}\left(E(\boldsymbol{P}^0_{t-1}|P_{t-1})|P_t\right) \\ & = & E\left(P_{t-1}(1-P_{t-1})|P_t\right)\boldsymbol{\Lambda}_{t-1} \\ & = & P_t(1-P_t)(1+O(N^{-1}_{eV,t-1}))\boldsymbol{\Lambda}_{t-1} \end{array} \tag{61}$$

using conditional independence of $\boldsymbol{P}^0_{t-1}$ and ε_t given P_{t-1} in the second step, (17) in the third step and (58)-(59) in the last step.

After some computations, it follows from (19) that

$$\mathrm{Cov}\left((\varepsilon_t, \boldsymbol{\varepsilon}^0_t)|P_{t-1}\right) = P_{t-1}(1-P_{t-1})\begin{pmatrix} \boldsymbol{\gamma}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T & \boldsymbol{\sigma}^T_{t-1} \\ \boldsymbol{\sigma}_{t-1} & \boldsymbol{\Sigma}^0_{t-1} \end{pmatrix}, \tag{62}$$

where

$$\boldsymbol{\sigma}_{t-1} := \boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T - (\boldsymbol{\gamma}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T)\mathbf{1}$$

and

$$\boldsymbol{\Sigma}^0_{t-1} := \boldsymbol{\Sigma}_{t-1} - (\boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T)\mathbf{1}^T - \mathbf{1}(\boldsymbol{\gamma}\boldsymbol{\Sigma}_{t-1}) + (\boldsymbol{\gamma}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T)\mathbf{1}\mathbf{1}^T. \tag{63}$$

By the multivariate central limit theorem, for large populations, we can justify that $(\varepsilon_t, \boldsymbol{\varepsilon}^0_t)|P_{t-1}$ approximately has a multivariate normal distribution, with covariance matrix as in (62). This implies

$$\begin{array}{rcl} E(\boldsymbol{\varepsilon}^0_t|P_{t-1},\varepsilon_t) & = & \boldsymbol{\sigma}_{t-1}\varepsilon_t/(\boldsymbol{\gamma}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T), \\ \mathrm{Cov}(\boldsymbol{\varepsilon}^0_t|P_{t-1},\varepsilon_t) & = & P_{t-1}(1-P_{t-1})\left(\boldsymbol{\Sigma}^0_{t-1} - \boldsymbol{\sigma}_{t-1}\boldsymbol{\sigma}^T_{t-1}/(\boldsymbol{\gamma}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T)\right), \end{array}$$

and hence

$$\begin{array}{rcl} \mathrm{Cov}(\boldsymbol{\varepsilon}^0_t|P_t) & = & E\left(\mathrm{Cov}(\boldsymbol{\varepsilon}^0_t|P_{t-1},\varepsilon_t)|P_t\right) + \mathrm{Cov}\left(E(\boldsymbol{\varepsilon}^0_t|P_{t-1},\varepsilon_t)|P_t\right) \\ & = & E\left(P_{t-1}(1-P_{t-1})|P_t\right)\left(\boldsymbol{\Sigma}^0_{t-1} - \boldsymbol{\sigma}_{t-1}\boldsymbol{\sigma}^T_{t-1}/(\boldsymbol{\gamma}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T)\right) \\ & + & \boldsymbol{\sigma}_{t-1}\boldsymbol{\sigma}^T_{t-1}\mathrm{Var}(\varepsilon_t|P_t)/(\boldsymbol{\gamma}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T)^2 \\ & = & P_t(1-P_t)\left(\boldsymbol{\Sigma}^0_{t-1} - \boldsymbol{\sigma}_{t-1}\boldsymbol{\sigma}^T_{t-1}/(\boldsymbol{\gamma}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T)\right) \\ & + & \boldsymbol{\sigma}_{t-1}\boldsymbol{\sigma}^T_{t-1}P_t(1-P_t)/(\boldsymbol{\gamma}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\gamma}^T) + O(N^{-1}_{eV,t-1}) \\ & = & P_t(1-P_t)\boldsymbol{\Sigma}^0_{t-1} + O(N^{-1}_{eV,t-1}), \end{array} \tag{64}$$

using (58)-(59) in the third equality. Combining (60), (61) and (64), we thus get

$$\boldsymbol{\Lambda}_t = \boldsymbol{B}^0\boldsymbol{\Lambda}_{t-1}(\boldsymbol{B}^0)^T + \boldsymbol{\Sigma}^0_{t-1} + R_t, \tag{65}$$

where $\boldsymbol{R}_t = O(N^{-1}_{eV,t-1})$ is a remainder term matrix. Repeated application of (65) gives

$$\boldsymbol{\Lambda}_t = \sum_{r=0}^{\infty}(\boldsymbol{B}^0)^r(\boldsymbol{\Sigma}^0_{t-r-1} + \boldsymbol{R}_{t-r})((\boldsymbol{B}^0)^r)^T, \tag{66}$$

which converges since all eigenvalues of $\boldsymbol{B}^0$ have modulus less than one. Moreover, since $|\lambda_2|$ is the maximal modulus of the eigenvalues of $\boldsymbol{B}^0$, it follows from (51), (57) and (66) that $\boldsymbol{\Sigma}_t \to \boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}_t \to \boldsymbol{\Lambda}$, where $\boldsymbol{\Sigma}$ satisfies the upper part of (22) and

$$\begin{aligned}\boldsymbol{\Lambda} &= \textstyle\sum_{r=0}^{\infty}(\boldsymbol{B}^0)^r\boldsymbol{\Sigma}^0((\boldsymbol{B}^0)^r)^T \\ &= \boldsymbol{V}\left(\textstyle\sum_{r=0}^{\infty}(\boldsymbol{D}^0)^r\boldsymbol{V}^{-1}\boldsymbol{\Sigma}^0\boldsymbol{V}^{-T}((\boldsymbol{D}^0)^r)^T\right)\boldsymbol{V}^T.\end{aligned} \tag{67}$$

In formula (67), $\boldsymbol{V}^T$ refers to transposition *and* complex conjugation and (54) is used in the second step. The matrix

$$\boldsymbol{\Sigma}^0 := \boldsymbol{\Sigma} - (\boldsymbol{\Sigma}\boldsymbol{\gamma}^T)\mathbf{1}^T - \mathbf{1}(\boldsymbol{\gamma}\boldsymbol{\Sigma}) + (\boldsymbol{\gamma}\boldsymbol{\Sigma}\boldsymbol{\gamma}^T)\mathbf{1}\mathbf{1}^T$$

is defined as in (63), but without index $t-1$ on the right hand side. To summarize, we have motivated the convergence in (18) and (19).

It remains to motivate the lower part of (22) by rewriting (67) in the appropriate way. To this end, it will be convenient to first find a square matrix $\boldsymbol{H}$ of order s^2 such that $\text{vec}(\boldsymbol{\Lambda}) = \boldsymbol{H}\text{vec}(\boldsymbol{\Sigma})$, where vec is the vectorization operator that transforms a square matrix of order s to a column vector of length s^2 by listing the columns of $\boldsymbol{\Lambda}$ from left to right. For simplicity of notation we write $\text{vec}(\boldsymbol{\Lambda})_{ij} = \Lambda_{ij}$ for component number $i + s(j-1)$ of $\text{vec}(\boldsymbol{\Lambda})$ and $\boldsymbol{H} = (H_{uv,ij})$, so that the componentwise relation between $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$ can be written as

$$\Lambda_{uv} = \sum_{i,j=1}^{s} H_{uv,ij}\Sigma_{ij}. \tag{68}$$

It follows from (67) and (63) that

$$\boldsymbol{H} = \boldsymbol{H}^{(4)}\boldsymbol{H}^{(3)}\boldsymbol{H}^{(2)}\boldsymbol{H}^{(1)}, \tag{69}$$

where $\text{vec}(\boldsymbol{\Sigma}^0) = \boldsymbol{H}^{(1)}\text{vec}(\boldsymbol{\Sigma})$, i.e.

$$H^{(1)}_{uv,ij} = 1_{\{(u,v)=(i,j)\}} - \gamma_j 1_{\{u=i\}} - \gamma_i 1_{\{v=j\}} + \gamma_i\gamma_j, \tag{70}$$

$\text{vec}(\boldsymbol{V}^{-1}\boldsymbol{X}\boldsymbol{V}^{-T}) = \boldsymbol{H}^{(2)}\text{vec}(\boldsymbol{X})$ and $\boldsymbol{V}^{-1} = (V_{ij}^{(-1)})$, i.e.

$$H_{uv,ij}^{(2)} = V_{ui}^{(-1)}V_{vj}^{(-1)}, \tag{71}$$

$\text{vec}\left(\sum_{r=0}^{\infty}(\boldsymbol{D}^0)^r\boldsymbol{X}((\boldsymbol{D}^0)^r)^T\right) = \boldsymbol{H}^{(3)}\boldsymbol{X}$, i.e.

$$H_{uv,ij}^{(3)} = \frac{1_{\{(u,v)=(i,j)\}}}{1 - D_{uu}^{(0)}D_{vv}^{(0)}} \tag{72}$$

in the special case when $\boldsymbol{D}$ is diagonal (otherwise $\boldsymbol{H}^{(3)}$ gets more complicated), and $\text{vec}(\boldsymbol{V}\boldsymbol{X}\boldsymbol{V}^T) = \boldsymbol{H}^{(4)}\text{vec}(\boldsymbol{X})$, i.e.

$$H_{uv,ij}^{(4)} = V_{ui}V_{vj}. \tag{73}$$

Analogously to (68), we write $\boldsymbol{G} = (G_{uv,ij})$ for $u \geq v$ and $i \geq j$, with

$$\Lambda_{uv} = \sum_{1\leq j\leq i\leq s} G_{uv,ij}\Sigma_{ij}. \tag{74}$$

Since $\boldsymbol{\Sigma}$ is symmetric, we may compute the elements of $\boldsymbol{G}$ from those of $\boldsymbol{H}$ according to

$$G_{uv,ij} = \begin{cases} H_{uv,ij} + H_{uv,ji}, & \text{if } i \neq j, \\ H_{uv,ii}, & \text{if } i = j. \end{cases} \tag{75}$$

Summarizing, the lower part of (22) is deduced from (68)-(75). $\square$

Motivating (20), (21) and (50). Assume that quasi equilibrium has been attained, so that $\boldsymbol{\Lambda}_t$ and $\boldsymbol{\Sigma}_t$ can be replaced by $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$ in (18) and (19) respectively. We start by motivating (20), using the approximation $E(Y/X) \approx E(Y)/E(X)$ to conclude that

$$N_{eV}^{\text{eq}} = E(N_{eV,t}) \approx \frac{E\left(p_t(1-p_t)\right)}{2E\left(E\left((p_{t+1}-p_t)^2|p_t\right)\right)}, \tag{76}$$

where expectation is with respect to the quasi equilibrium distribution. For the numerator of (76), we first condition on P_t, and notice that

$$\begin{aligned} E\left(p_t(1-p_t)|P_t\right) &= P_t(1-P_t) - E\left((p_t-P_t)^2|P_t\right) \\ &= P_t(1-P_t) - E\left(((\boldsymbol{a}-\boldsymbol{\gamma})\boldsymbol{P}_t^0)^2|P_t\right) \\ &= P_t(1-P_t)\left(1 - (\boldsymbol{a}-\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{a}-\boldsymbol{\gamma})^T\right), \end{aligned} \tag{77}$$

where in the second equality we used

$$p_t - P_t = (\boldsymbol{a}-\boldsymbol{\gamma})\boldsymbol{P}_t = (\boldsymbol{a}-\boldsymbol{\gamma})\boldsymbol{P}_t^0 = \boldsymbol{a}\boldsymbol{P}_t^0.$$

Then averaging (77) with respect to P_t we get

$$\begin{aligned} E\left(p_t(1-p_t)\right) &= E\left(E\left(p_t(1-p_t)|P_t\right)\right) \\ &= E\left(P_t(1-P_t)\right)\left(1-(\boldsymbol{a}-\boldsymbol{\gamma})\boldsymbol{\Lambda}(\boldsymbol{a}-\boldsymbol{\gamma})^T\right). \end{aligned} \tag{78}$$

For the denominator of (76) we use (14) and $(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{1}=\boldsymbol{0}$ to deduce that

$$\begin{aligned} p_{t+1} &= \boldsymbol{a}\boldsymbol{P}_{t+1} \\ &= \boldsymbol{a}\boldsymbol{B}\boldsymbol{P}_t + \boldsymbol{a}\boldsymbol{\varepsilon}_{t+1} \\ &= p_t + \boldsymbol{a}(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{P}_t + \boldsymbol{a}\boldsymbol{\varepsilon}_{t+1} \\ &= p_t + \boldsymbol{a}(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{P}_t^0 + \boldsymbol{a}\boldsymbol{\varepsilon}_{t+1}. \end{aligned}$$

Hence

$$\begin{aligned} & E\left(E\left((p_{t+1}-p_t)^2|p_t\right)\right) = E\left(E\left((p_{t+1}-p_t)^2|p_t,P_t\right)\right) \\ &\quad = \boldsymbol{a}(\boldsymbol{B}-\boldsymbol{I})E\left(E\left(\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T|p_t,P_t\right)\right)(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{a}^T + \boldsymbol{a}E\left(E\left(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}_{t+1}^T|p_t,P_t\right)\right)\boldsymbol{a}^T \\ &\quad = \boldsymbol{a}(\boldsymbol{B}-\boldsymbol{I})E\left(E\left(\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T|\boldsymbol{a}\boldsymbol{P}_t^0 = p_t - P_t\right)\right)(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{a}^T + \boldsymbol{a}E\left(E\left(\boldsymbol{\varepsilon}_{t+1}\boldsymbol{\varepsilon}_{t+1}^T|P_t\right)\right)\boldsymbol{a}^T \\ &\quad = \boldsymbol{a}(\boldsymbol{B}-\boldsymbol{I})E\left(\boldsymbol{P}_t^0(\boldsymbol{P}_t^0)^T\right)(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{a}^T + E\left(P_t(1-P_t)\right)\boldsymbol{a}\boldsymbol{\Sigma}\boldsymbol{a}^T \\ &\quad = E\left(P_t(1-P_t)\right)\left(\boldsymbol{a}(\boldsymbol{B}-\boldsymbol{I})\boldsymbol{\Lambda}(\boldsymbol{B}-\boldsymbol{I})^T\boldsymbol{a}^T + \boldsymbol{a}\boldsymbol{\Sigma}\boldsymbol{a}^T\right), \end{aligned}$$

where the outer expectation is with the respect to p_t *and* P_t under the quasi equilibrium distribution. Taking the ratio of (78) and the last equation and inserting into (76), we arrive at (20).

In order to verify (21), we we notice that

$$F_{ST}^{\text{eq}} = E(F_{ST,t}) \approx \frac{\sum_{i=1}^s a_i E\left((P_{ti}-p_t)^2\right)}{E\left(p_t(1-p_t)\right)}, \tag{79}$$

using again the approximation $E(Y/X) \approx E(Y)/E(X)$ in the second step. In order to further expand the right hand side of (79), we write

$$\boldsymbol{P}_t - p_t\boldsymbol{1} = (\boldsymbol{I}-\boldsymbol{1}\boldsymbol{a})\boldsymbol{P}_t = (\boldsymbol{I}-\boldsymbol{1}\boldsymbol{a})\boldsymbol{P}_t^0,$$

which leads to

$$\begin{aligned} E\left((P_{ti}-p_t)^2\right) &= E\left(E\left((P_{ti}-p_t)^2|P_t\right)\right) \\ &= E\left(P_t(1-P_t)\right)\left((\boldsymbol{I}-\boldsymbol{1}\boldsymbol{a})\boldsymbol{\Lambda}(\boldsymbol{I}-\boldsymbol{1}\boldsymbol{a})^T\right)_{ii}. \end{aligned} \tag{80}$$

Inserting (78) and (80) into (79) and dividing the numerator and denominator of (79) by $E\left(P_t(1-P_t)\right)$, we arrive at (21).

In order to verify (50), we write

$$G_{ST,t} \approx \frac{\sum_{lq}\sum_{i=1}^s a_i E\left(P_{ti}^{lq}-p_t^{lq})^2\right)}{\sum_{lq} E\left(p_t^{lq}(1-p_t^{lq})\right)} = \sum_{lq}\omega_{lq}\frac{\sum_{i=1}^s a_i E\left(P_{ti}^{lq}-p_t^{lq})^2\right)}{E\left(p_t^{lq}(1-p_t^{lq})\right)}. \tag{81}$$

The approximation $E(Y/X) \approx E(Y)/E(X)$ is increasingly accurate the larger L is. Indeed, as L grows, both X and Y get more concentrated around their expected values, with coefficients of variation inversely proportional to $\sqrt{L}$ when the loci are in linkage equilibrium. Notice that the right hand side of (81) is written as a weighted average of terms of the same type as used for the approximation in (79), with weights ω_{lq} proportional to $E\left(p_t^{lq}(1-p_t^{lq})\right)$ summing to one. Applying (21) to each of these terms, we deduce (50).

In an analogous manner, the approximation in (20) for the effective population size can be motivated for multiple loci and markers by first rewriting (12) as

$$N_{eV} = \frac{\sum_{lq} p_t^{lq}(1-p_t^{lq})}{\sum_{lq} 2E\left((p_{t+1}^{lq}-p_t^{lq})^2|p_t^{lq}\right)} \approx \frac{\sum_{lq} E\left(p_t^{lq}(1-p_t^{lq})\right)}{\sum_{lq} 2E\left(E\left((p_{t+1}^{lq}-p_t^{lq})^2|p_t^{lq}\right)\right)}, \tag{82}$$

where in the last step we use a coefficient of variation argument separately for the numerator and denominator. The right hand side of (82) can be written as a weighted average of terms identical to the right hand side of (76), which in turn equals (20). □

Reproduction scenario 2. We will verify (55) (and hence also (56)) separately for reproduction scenarios 2, 1 and 3. Starting with reproduction scenario 2, we write

$$P_{tki}^* = P_{tk} + (\tilde{P}_{tk} - P_{tk}) + (P_{tki}^* - \tilde{P}_{tk}).$$

It follows from (9) and (24) that

$$\varepsilon_{t+1,i} = \sum_{k=1}^{s} b_{ik}(\tilde{P}_{tk} - P_{tk}) + \sum_{k=1}^{s} b_{ik}(P_{tki}^* - \tilde{P}_{tk}).$$

We further have that

$$\text{Var}(\tilde{P}_{tk} - P_{tk}|\boldsymbol{P}_t) = P_{tk}(1-P_{tk})\left(\frac{1}{2N_{ek}} - \frac{1}{2Na_k}\right)(1+o(1)) \tag{83}$$

and

$$\text{Var}(P_{tki}^* - \tilde{P}_{tk}|\boldsymbol{P}_t) = \frac{P_{tk}(1-P_{tk})}{2Na_km_{ki}}(1+o(1)).$$

From this it follows that

$$\boldsymbol{\Sigma}(\boldsymbol{P}_t)_{ij} = \sum_{k=1}^{s} b_{ik}b_{jk}P_{tk}(1-P_{tk})\left(\frac{1}{2N_{ek}} - \frac{1}{2Na_k}\right) + 1_{\{i=j\}}\sum_{k=1}^{s} b_{ik}^2\frac{P_{tk}(1-P_{tk})}{2Na_km_{ki}}.$$

Conditioning on P_t, taking expectation with respect to $\boldsymbol{P}_t^0$ (cf. (19)) and using the fact that $b_{ik}^2/(2Na_k m_{ki}) = b_{ik}/(2Na_i)$, we arrive at

$$\begin{array}{rcl} \Sigma_{ij} & = & \left(\sum_{k=1}^s b_{ik}b_{jk}\left(\frac{1}{2N_{ek}} - \frac{1}{2Na_k}\right) + \frac{1_{\{i=j\}}}{2Na_i}\right) \\ & - & \sum_{k=1}^s b_{ik}b_{jk}\left(\frac{1}{2N_{ek}} - \frac{1}{2Na_k}\right)\Lambda_{kk} - \frac{1_{\{i=j\}}}{2Na_i}\sum_{k=1}^s b_{ik}\Lambda_{kk} \\ & = & f_{ij} - \sum_{k=1}^s F_{ij,kk}\Lambda_{kk}, \end{array} \tag{84}$$

which verifies (56) for the breeders example. ☐

Reproduction scenario 1. Turning to the more general case (23), we write

$$\varepsilon_{t+1,i} = \sum_{k=1}^s b_{ik}(P_{tki}^* - P_{tk}). \tag{85}$$

Introduce $C_{kij} = \mathrm{Cov}(\nu_{ki}^l, \nu_{kj}^{l'})$ when $l \neq l'$. Because of the assumed exchangeability of $\{\boldsymbol{\nu}_k^l\}_{l=1}^{2Na_k}$, C_{kij} does not depend on (l, l'). Since (3) holds exactly, with remainder term $o(1)$ equal to zero, the variance of the left hand side must be zero, and this implies $C_{kij} = -V_{kij}/(2Na_k - 1)$. Therefore, it follows from (23) that

$$\begin{array}{rcl} \mathrm{Cov}(P_{tki}^*, P_{tkj}^* | \boldsymbol{P}_t) & = & \frac{2Na_kP_{tk}V_{kij} + 2Na_kP_{tk}(2Na_kP_{tk}-1)C_{kij}}{(2Na_k)^2 m_{ki}m_{kj}} \\ & \sim & \frac{V_{kij}}{m_{ki}m_{kj}}\frac{P_{tk}(1-P_{tk})}{2Na_k}. \end{array}$$

Combining this with (85), we can first compute the conditional covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{P}_t)$, then condition on P_t and take expectation with respect to $\boldsymbol{P}_t^0$, as above, to deduce

$$\begin{array}{rcl} \Sigma_{ij} & = & \sum_{k=1}^s b_{ik}b_{jk}\frac{V_{kij}}{m_{ki}m_{kj}}\frac{1}{2Na_k} - \sum_{k=1}^s b_{ik}b_{jk}\frac{V_{kij}}{m_{ki}m_{kj}}\frac{1}{2Na_k}\Lambda_{kk} \\ & = & \sum_{k=1}^s \frac{V_{kij}a_k}{a_ia_j}\frac{1}{2N} - \sum_{k=1}^s \frac{V_{kij}a_k}{a_ia_j}\frac{1}{2N}\Lambda_{kk} \\ & = & f_{ij} - \sum_{k=1}^s F_{ij,kk}\Lambda_{kk}. \end{array} \tag{86}$$

We find that (84) and (86) agree if V_{kij} is chosen as in (25). ☐

Reproduction scenario 3. In order to verify (56), we first notice from (9) and (28) that

$$\varepsilon_{t+1,i} = (P_{t+1,i} - \check{P}_{ti}) + \sum_{k=1}^s b_{ik}(\tilde{P}_{tk} - P_{tk}) + \sum_{k=1}^s (B_{ik} - b_{ik})P_{tk} + \mathrm{rem}, \tag{87}$$

with $\mathrm{rem} = \sum_{k=1}^s (B_{ik} - b_{ik})(\tilde{P}_{tk} - P_{tk})$ a remainder term that vanishes when $N_{ek} = Na_k$ for all k and which is otherwise asymptotically negligible when $\alpha_i \to \infty$ as $N \to \infty$. It follows from (28) and (31) that

$$\mathrm{Var}(P_{t+1,i} - \check{P}_{ti} | \boldsymbol{P}_t) \sim \frac{(\boldsymbol{BP}_t)_i(1 - (\boldsymbol{BP}_t)_i)}{2Na_i}(1 + o(1)).$$

and

$$\begin{aligned} \operatorname{Var}\left(\sum_{k=1}^{s}(B_{ik}-b_{ik})P_{tk}|\boldsymbol{P}_t\right) &= \tfrac{1}{\alpha_i+1}\sum_{k=1}^{s}P_{tk}^2 b_{ik} - \tfrac{1}{\alpha_i+1}\sum_{j,k=1}^{s}P_{tj}P_{tk}b_{ij}b_{ik} \\ &= \tfrac{1}{\alpha_i+1}\sum_{k=1}^{s}\left(P_{tk}-(\boldsymbol{B}\boldsymbol{P}_t)_i\right)^2 b_{ik}. \end{aligned}$$

In conjunction with (83) and (87), this implies

$$\begin{aligned} \boldsymbol{\Sigma}(\boldsymbol{P})_{ij} &= \tfrac{1_{\{i=j\}}}{2Na_i}(\boldsymbol{B}\boldsymbol{P}_t)_i(1-(\boldsymbol{B}\boldsymbol{P}_t)_i) \\ &+ \sum_{k=1}^{s} b_{ik}b_{jk}P_{tk}(1-P_{tk})\left(\tfrac{1}{2N_{ek}}-\tfrac{1}{2Na_k}\right) \\ &+ \tfrac{1_{\{i=j\}}}{\alpha_i+1}\sum_{k=1}^{s}\left(P_{tk}-(\boldsymbol{B}\boldsymbol{P}_t)_i\right)^2 b_{ik}. \end{aligned}$$

Conditioning on P_t and taking expectation with respect to $\boldsymbol{P}_t^0$, we obtain

$$\begin{aligned} \Sigma_{ij} &= \sum_{k=1}^{s} b_{ik}b_{jk}\left(\tfrac{1}{2N_{ek}}-\tfrac{1}{2Na_k}\right) + \tfrac{1_{\{i=j\}}}{2Na_i} \\ &- \sum_{k=1}^{s} b_{ik}b_{jk}\left(\tfrac{1}{2N_{ek}}-\tfrac{1}{2Na_k}\right)\Lambda_{kk} - \tfrac{1_{\{i=j\}}}{2Na_i}(\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T)_{ii} \\ &+ \tfrac{1_{\{i=j\}}}{\alpha_i+1}\left(\sum_{k=1}^{s} b_{ik}\Lambda_{kk} - (\boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^T)_{ii}\right) \\ &= f_{ij} - \sum_{1\le v\le u} F_{ij,uv}\Lambda_{uv} \end{aligned} \tag{88}$$

after some computations, thus verifying (56). □

Deriving explicit expressions of $N_{eV}^{\mathbf{eq}}$ and $F_{ST}^{\mathbf{eq}}$ for the island model. Since $\boldsymbol{\gamma} = \boldsymbol{a}$ for the island model, we can apply (33) and (34), with $\boldsymbol{a} = \mathbf{1}^T/s$, to deduce

$$N_{eV}^{\mathrm{eq}} = \frac{1}{2\mathbf{1}^T\boldsymbol{\Sigma}\mathbf{1}/s^2} \tag{89}$$

and

$$F_{ST}^{\mathrm{eq}} = \frac{1}{s}\mathrm{tr}(\boldsymbol{\Lambda}). \tag{90}$$

We will start by giving a more explicit expression for $\boldsymbol{\Lambda}$. It follows from (38) that $\boldsymbol{B}\boldsymbol{v} = (1-m)\boldsymbol{v}$ for any vector $\boldsymbol{v}$ with $(\boldsymbol{v}, \mathbf{1}) = 0$. Hence $\lambda_2 = \ldots = \lambda_s = 1-m$. In this case it is particularly convenient to put $\lambda_1^0 = 1-m$ in the definition of $\boldsymbol{D}_0$, since then, according to (54), $\boldsymbol{B}^0 = (1-m)\boldsymbol{I}$. It then follows from (67) that

$$\boldsymbol{\Lambda} = \sum_{r=0}^{\infty}(1-m)^{2r}\boldsymbol{\Sigma}^0 = \frac{\boldsymbol{\Sigma}^0}{1-(1-m)^2}, \tag{91}$$

with $\boldsymbol{\Sigma}^0$ as in (63). Consequently, (90) can be rewritten as

$$\left(1-(1-m)^2\right)F_{ST}^{\mathrm{eq}} = \frac{1}{s}\mathrm{tr}(\boldsymbol{\Sigma}_0) = \frac{1}{s}\left(\mathrm{tr}(\boldsymbol{\Sigma}) - \frac{1}{s}\mathbf{1}^T\boldsymbol{\Sigma}\mathbf{1}\right). \tag{92}$$

Hence, in view of (89), (90) and (92), it suffices to find $\boldsymbol{\Sigma}$.

For reproduction scenario 2, formula (84) simplifies to

$$\begin{aligned}\Sigma_{ij} \;&=\; \left(\frac{1}{2N_e}-\frac{1}{2N/s}\right)\left(\frac{2m-m^2}{s}+(1-m)^2 1_{\{i=j\}}\right)+\frac{1_{\{i=j\}}}{2N/s}\\ &-\; \left(\frac{1}{2N_e}-\frac{1}{2N/s}\right)\left(\frac{m^2}{s^2}\mathrm{tr}(\mathbf{\Lambda})+\frac{\Lambda_{ii}+\Lambda_{jj}}{2}\left(2\frac{m}{s}(1-m)+1_{\{i=j\}}(1-m)^2\right)\right)\\ &-\; \frac{1_{\{i=j\}}}{2N/s}\left(\frac{m}{s}\mathrm{tr}(\mathbf{\Lambda})+(1-m)\Lambda_{ii}\right)\end{aligned} \tag{93}$$

for the island model, so that

$$\frac{2}{s^2}\mathbf{1}^T\mathbf{\Sigma}\mathbf{1}=\frac{1}{sN_e}\left(1-\frac{1}{s}\mathrm{tr}(\mathbf{\Lambda})\right)=\frac{1}{sN_e}(1-F_{ST}^{\mathrm{eq}}) \tag{94}$$

and

$$\frac{1}{s}\mathrm{tr}(\mathbf{\Sigma}^0)=\frac{s-1}{s}\frac{1}{2\tilde{N}}(1-F_{ST}^{\mathrm{eq}}). \tag{95}$$

Combining (89) and (94) we arrive at (39), and inserting (95) into (92) and solving for F_{ST}^{eq} we arrive at (40).

For reproduction scenario 3, a similar simplification of (88) leads to

$$\begin{aligned}\frac{2}{s^2}\mathbf{1}^T\mathbf{\Sigma}\mathbf{1} \;&=\; \frac{1}{sN_e}-\left(\frac{1}{N_e}-\frac{1-(1-m)^2}{N/s}\right)\frac{1}{s^2}\mathrm{tr}(\mathbf{\Lambda})+\frac{2\big(1-(1-m)^2\big)}{\alpha+1}\frac{1}{s^2}\mathrm{tr}(\mathbf{\Lambda})\\ &=\; \frac{1}{sN_e}-\left(\frac{1}{N_e}-\frac{1-(1-m)^2}{N/s}\right)\frac{1}{s}F_{ST}^{\mathrm{eq}}+\frac{2\big(1-(1-m)^2\big)}{\alpha+1}\frac{1}{s}F_{ST}^{\mathrm{eq}}.\end{aligned} \tag{96}$$

and

$$\begin{aligned}\frac{1}{s}\mathrm{tr}(\mathbf{\Sigma}^0) \;&=\; \frac{s-1}{s}\left(\frac{1}{2\tilde{N}}-\frac{(1-m)^2}{2N_e}\frac{1}{s}\mathrm{tr}(\mathbf{\Lambda})+\frac{1-(1-m)^2}{\alpha+1}\frac{1}{s}\mathrm{tr}(\mathbf{\Lambda})\right)\\ &=\; \frac{s-1}{s}\left(\frac{1}{2\tilde{N}}-\frac{(1-m)^2}{2N_e}F_{ST}^{\mathrm{eq}}+\frac{1-(1-m)^2}{\alpha+1}F_{ST}^{\mathrm{eq}}\right).\end{aligned} \tag{97}$$

Inserting (96) into (89) we arrive at (42), and plugging (97) into (92) and solving for F_{ST}^{eq} we arrive at (43). □

Acknowledgements

Ola Hössjer's research was financially supported by the Swedish Research Council, contract nr. 621-2008-4946, and the Gustafsson Foundation for Research in Natural Sciences and Medicine. Nils Ryman's research was supported by grants from the Swedish Research Council, the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas), and from the BONUS Baltic Organisations' Network for Funding Science EEIG (the BaltGene research project).

References

Allendorf, F.W. and Luikart, G. (2007). *Conservation and the genetics of populations.* Blackwell Publishing, Malden, MA.

Brockwell, P.J. and Davis, R.A. (1987). *Time Series: Theory and Methods*, Springer-Verlag, New York.

Caballero, A. (1994). Developments in the prediction of effective population size. *Heredity* **73**, 657-679.

Caswell, H. (2001). *Matrix Population Models*, 2nd ed., Sinauer, Sunderland, MA.

Cox, D.R. and Miller, H.D. (1965). *The theory of Stochastic Processes.* Methuen & Co Ltd, London.

Durrett, R. (2008). *Probability models for DNA sequence evolution*, 2nd ed., Springer, New York.

Engen, S., Lande, R. and Saether, B-E. (2005a). Effective size of a fluctuating age-structured population. *Genetics* **170**, 941-954.

Engen, S., Lande, R., Saether, B-E. and Weimerskirch, H. (2005b). Extinction in relation to demographic and environmental stochasticity in age-structured models. *Mathematical Biosciences* **195**, 210-227.

Engle, R.F. and Granger, C.W.J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica* **55**, 251-276.

Ewens, W.J. (2004). *Mathematical Population Genetics.* I. Theoretical introduction, 2nd ed., Springer, New York.

Granger, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* **16**, 121-130.

Hardy, O.J. and Vekemans, X. (2002). SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population model. *Molecular Ecology Notes* **2**, 618-620.

Hare, M.P., Nunney, L., Schwartz, M.K., Ruzzante, D.E., Burford, M., Waples, R., Ruegg, K. and Palstra, F. (2011). Understanding and estimating effective population size for practical applications in marine species management. *Conservational Biology* **25**(3), 438-449.

Hössjer, O. (2012). Coalescence Theory for a General Class of Structured Populations with Fast Migration. *Advances of applied probability* **43**(4), 1027-1047.

Hössjer, O., Jorde, P.E. and Ryman, N. (2012). Quasi Equilibrium Approximations of the Fixation Index of the Island Model under Neutrality. Submitted.

Kimura, M. (1953). 'Stepping stone' model of population. *Ann. Rep. Natl. Inst. Genet. Japan* **3**, 62-63.

Kimura, M. and Weiss, G.H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **61**, 763-771.

Maruyama, T. (1970a). On the rate of decrease of heterozygosity in circular stepping stone models of populations. *Theor. Pop. Biol.* **1**, 101-119.

Maruyama, T. (1970b). Effective number of alleles in subdivided populations. *Theor. Pop. Biol.* **1**, 273-306.

Nagylaki, T. (1980). The strong migration limit in geographically structured populations. *J. Math. Biol.* **9**, 101-114.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* **70**, 3321-3323.

Nei, M. and Kumar, S. (2000). *Molecular evolution and phylogenetics.* Oxford University Press, Oxford.

Nordborg, M. and Krone, S. (2002). Separation of time scales and convergence to the coalescent in structured populations. In *Modern Development in Theoretical Population Genetics*, Slatkin, M. and Veuille, M., eds., 194-232. Oxford Univ Press.

Nunney, L. (1999). The effective size of a hierarchically-structured population. *Evolution* **53**, 1-10.

Orrive, M.E. (1993). Effective population size in organisms with complex life-histories. *Theor. Popul. Biol.* **44**, 316-340.

Palstra, F.P. and Ruzzante, D.E. (2008). Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild populations persistence? *Molecular Ecology* **17**, 3428-3447.

Ryman, N., Hössjer, O., Allendorf, F.W., Jorde P.E. and Laikre, L. (2012). Samples from structured populations yield biased estimates of effective size that overestimate the rate of loss of genetic variation. Submitted.

Ryman, N. and Leimar, O. (2008). Effect of mutation on genetic differentiation among nonequilibrium populations. *Evolution* **62**(9), 2250-2259.

Sagitov, S. and Jagers. P. (2005). The coalescent effective size of age-structured populations. *Ann. Appl. Probab.* **15**(3), 1778-1797.

Sved, J.A. and Latter, B.D.H. (1977). Migration and mutation in stochastic models of gene frequency change. *J. Math. Biology* **5**, 61-73.

Wang, J. and Caballero, A. (1999). Developments in predicting the effective size of subdivided populations. *Heredity* **82**, 212-226.

Waples, R.S. (2002). Definition and estimation of effective population size in the conservation of endangered species. In *Populations viability analysis*, eds. Beissinger, S.R. and McCullogh, D.R. The University of Chicago Press, Chicago.

Waples R.S. and Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.

Weiss, G.H. and Kimura, M. (1965). A mathematical analysis of the stepping stone model of genetic correlation. *J. Appl. Prob.* **2**, 129-149.

Whitlock, M.C. and Barton, N.H. (1997). The effective size of a subdivided population. *Genetics* **145**, 427-441.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97-159.

Wright, S. (1938). Size of population and breeding structure in relation to evolution. *Science* **87**, 430-431.

Wright, S. (1943). Isolation by distance. *Genetics* **28**, 114-138.

Wright, S. (1946). Isolation by distance under diverse systems of mating. *Genetics* **31**, 39-59.

Wright, S. (1951). The general structure of populations. *Annals of Eugenics* **15**, 323-354.

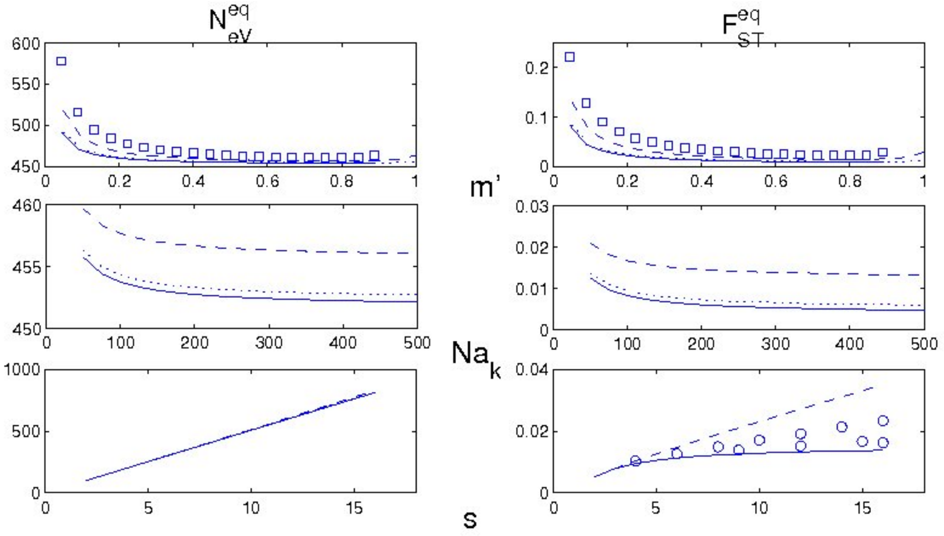

Figure 1: Plots of variance effective population size N_{eV}^{eq} (left) and fixation index F_{ST}^{eq} (right) when fertilization precedes migration for the island model (solid), circular (dashed), linear (squares) and torus (dotted and circles) stepping stone models. Upper: The migration rate m' in (6) is varied, whereas the number of subpopulations $s = 9$, the total population size $N = 450$, the relative subpopulation sizes $a_k = 1/9$, local effective population sizes $N_{ek} = N_e = N/9 = 50$ and (for the torus model) the grid sizes are $s_1 = s_2 = 3$. Middle: Local census size $Na_k = N/9$ is varied, whereas $s = 9$, $m' = 0.4$ and $N_{ek} = N_e = 50$. Lower: s is varied, whereas $N = 450$, $m' = 0.4$ and $N_{ek} = N_e = Na_k = 50$. For the torus stepping stone model (circles), the upper circles correspond to $s_1 = 2, \ldots, 8$ and $s_2 = 2$, the middle circles to $s_1 = 3, 4, 5$ and $s_2 = 3$, and the lower right circle to $s_1 = s_2 = 4$.

Table 1: **Notation used in the paper.**

Symbol	Definition
s	Number of subpopulations.
N	Total census size.
a_i	Relative size of subpopulation i.
Na_i	Local census size of subpopulation i.
ν_{ki}^l	Nr. of offspring in subpopulation i of l:th gene of subpopulation k.
m_{ki}	Migration rate from subpopulation k to i.
m'	Overall migration rate.
b_{ik}	Backward transition probability from subpopulation i to k.
γ_i	Equilibrium prob. of distant ancestor to come from subpop. i.
$\lvert\mathcal{N}\rvert$	Average relative size of one step neighborhood.
N_{ei}	Local effective size of subpopulation i.
N_e	Local effective size of all subpopulations (if constant).
$N_{eV,t}$	Variance effective size of population in generation t.
N_{eV}^{eq}	Quasi equilibrium value of N_{eV}.
$F_{ST,t}$	Fixation index of population in generation t.
F_{ST}^{eq}	Quasi equilibrium value of F_{ST}.
P_{ti}	Allele frequency of subpopulation i and generation t.
p_t	Overall allele frequency (weights a_i) in generation t.
P_t	Overall allele frequency (weights γ_i) in generation t.
ε_{ti}	Random drift of allele frequency, subpopulation i and generation t.
$\mathbf{\Sigma}$	Standardized cov. matrix of random drift of allele frequencies.
$\mathbf{\Lambda}$	Standardized cov. matrix of spatial allele frequency fluctuations.

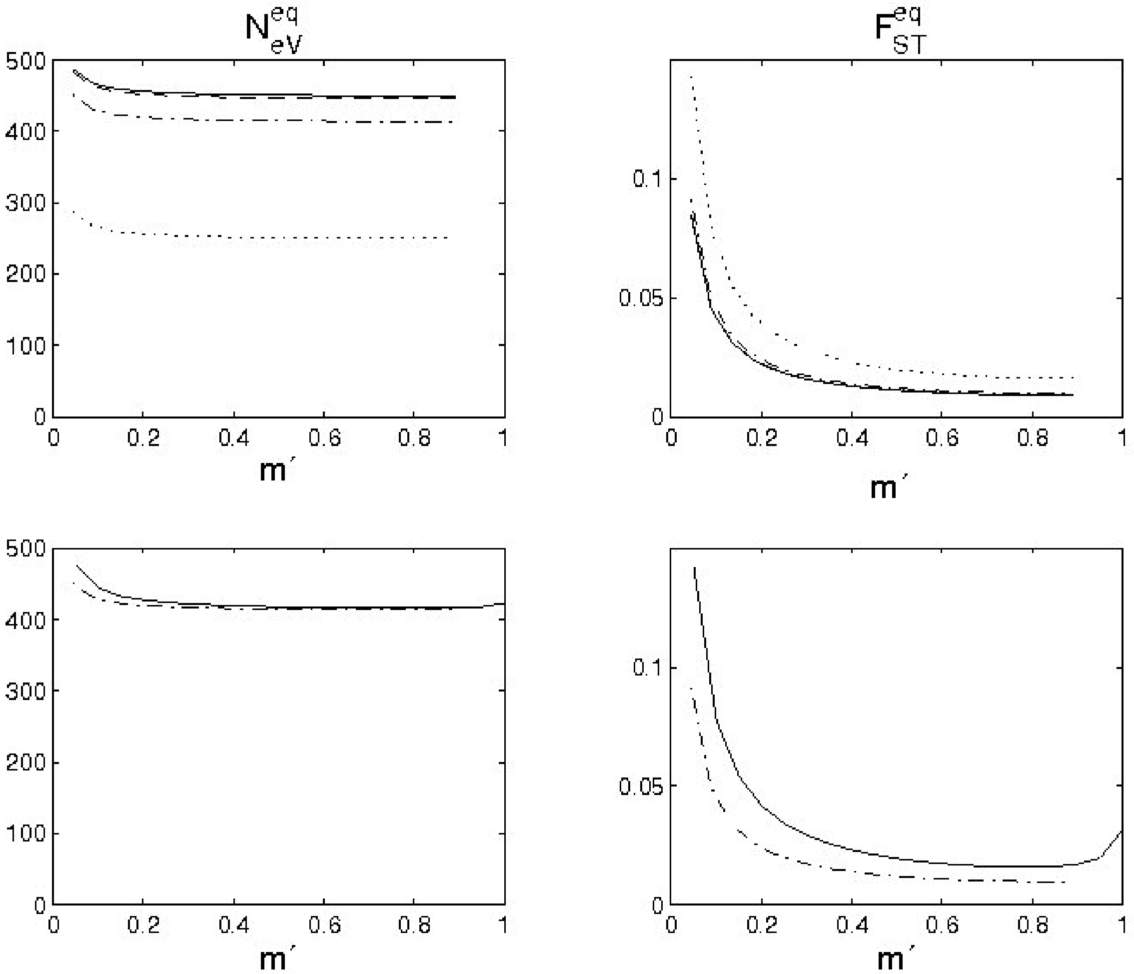

Figure 2: Plots of N_{eV}^{eq} (left) and F_{ST}^{eq} (right) as functions of m' when migration precedes fertilization with Dirichlet parameters $\alpha_k = \alpha$ (cf. (30)). For all curves $s = 9$, $N = 450$, $a_k = 1/9$ and $N_{ek} = N_e = N/9 = 50$. Upper: Plots for the island model with $\alpha = \infty$ (solid), $\alpha = 100$ (dashed), $\alpha = 10$ (dash-dotted) and $\alpha = 1$ (dotted). Lower: Plots, with $\alpha = 10$, for the island (dash-dotted) and circular stepping stone models (solid).

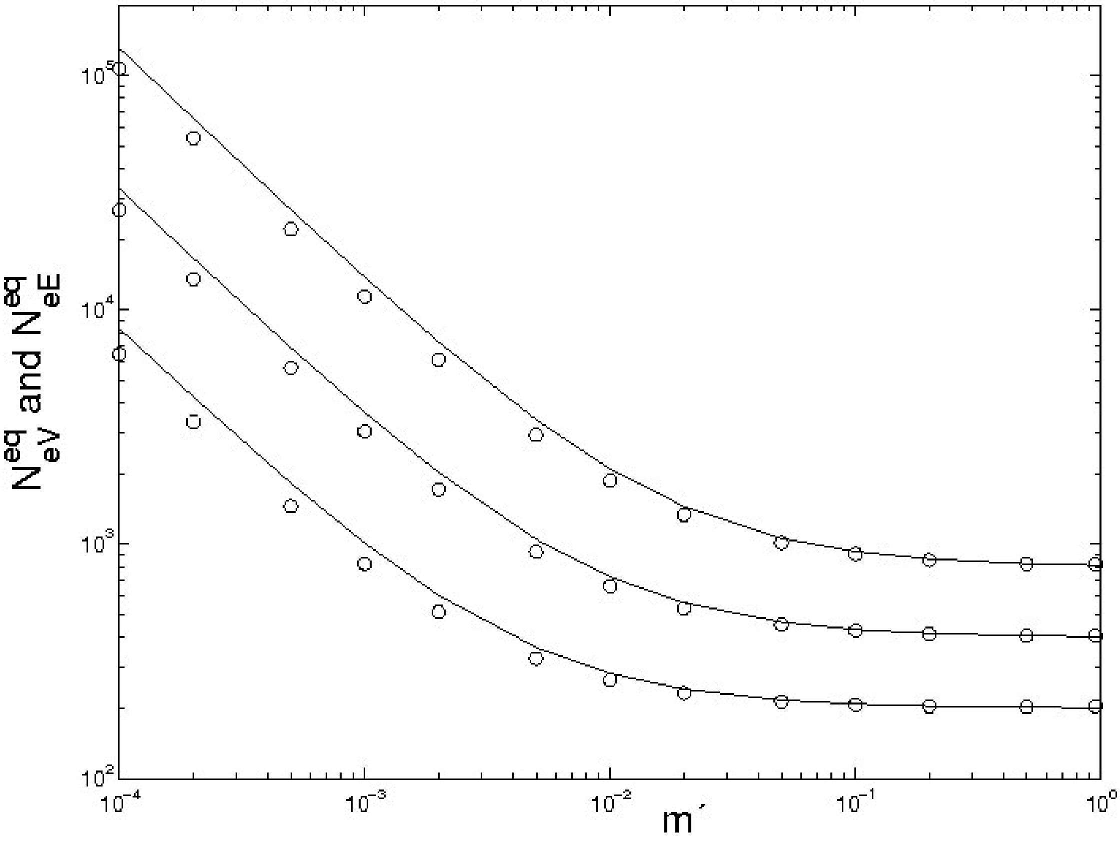

Figure 3: Plots of variance and eigenvalue effective population sizes N_{eV}^{eq} (dots) and N_{eE}^{eq} (solid lines, cf. (45)), as functions of the migration rate m' for the circular stepping stone model. The plots are on a log-log scale and the number of subpopulations is $s = 16$ (upper), $s = 8$ (middle) and $s = 4$ (lower). Fertilization precedes migration, with population size $N = 50s$. For all subpopulations k, $a_k = 1/s$ and $N_{ek} = N_e = Na_k = 50$.

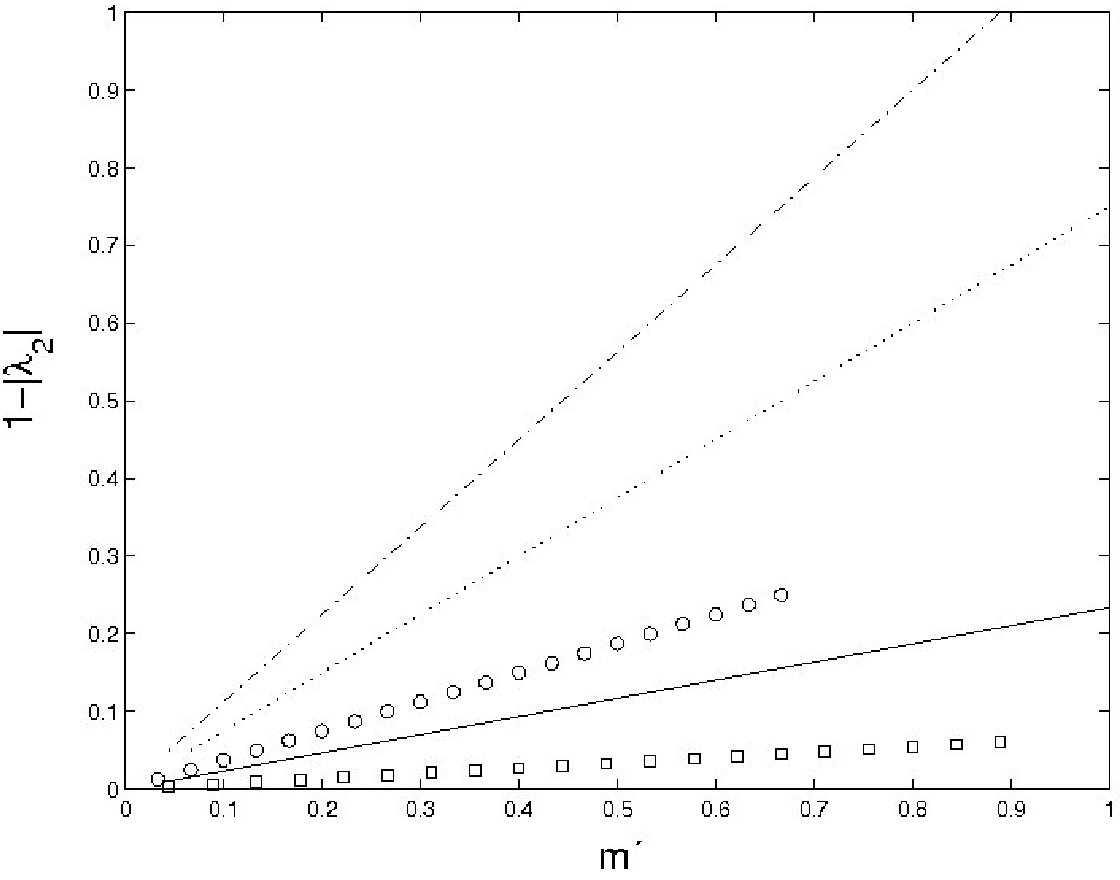

Figure 4: Plots of $1 - |\lambda_2|$ as function of the migration rate m' for a number of models with $s = 9$, $N = 450$, $a_k = 1/9$, $N_e = N_{ek} = N/9 = 50$ and different average one step neighbourhood sizes $|\mathcal{N}|$: The island model (dash-dotted, $|\mathcal{N}| = 1$), the torus stepping stone model (dotted, $|\mathcal{N}| = 1/2$), the rectangular stepping stone model (circles, $|\mathcal{N}| = 1/3$), the circular stepping stone model (solid, $|\mathcal{N}| = 1/4$) and the linear stepping stone model (squares, $|\mathcal{N}| = 2/9$). For the two-dimensional (rectangular and torus) stepping stone models $s_1 = s_2 = 3$.

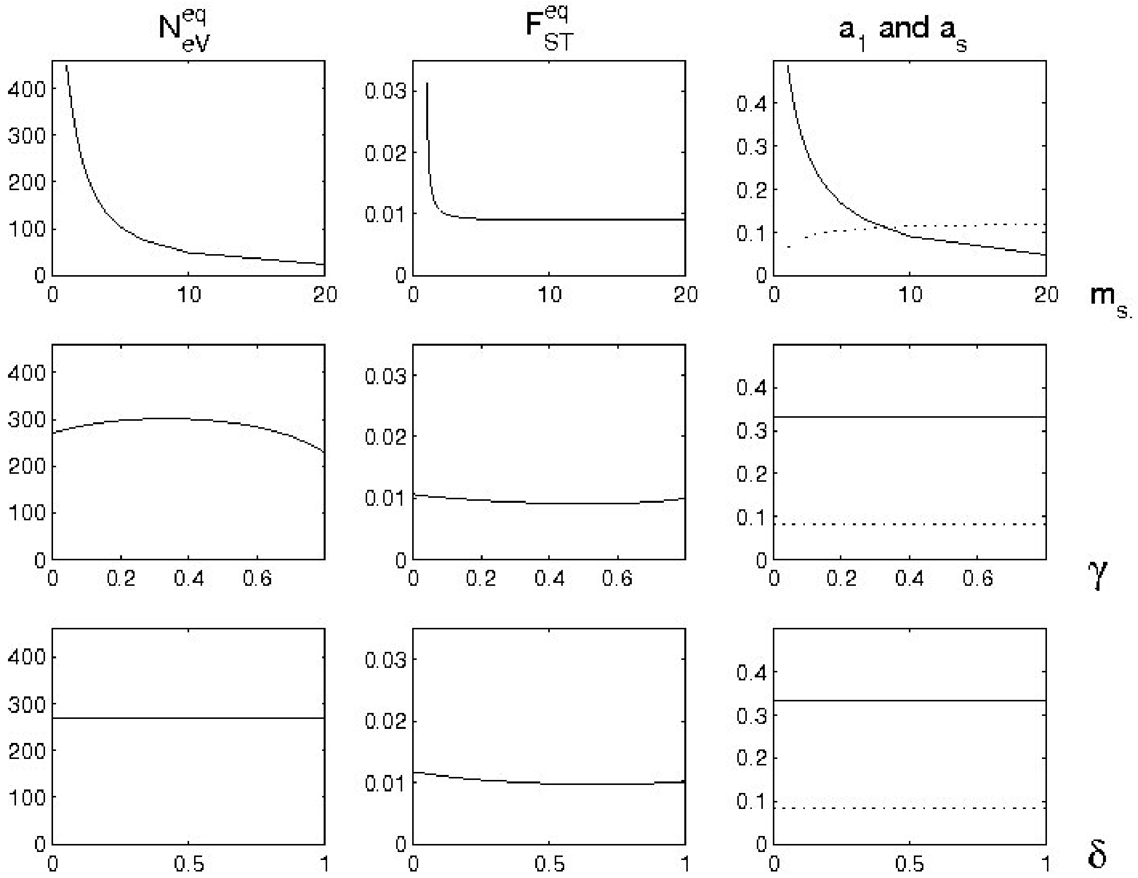

Figure 5: Plots, for the genetic reservoir model, of N_{eV}^{eq} (left), F_{ST}^{eq} (middle) and relative sizes (right) $a_1 = \ldots = a_{s-1}$ for the $s-1$ perimeter populations (dotted) and a_s for the reservoir population (solid), when fertilization precedes migration, $s = 9$, $N = 450$ and $N_{ek} = Na_k$ for $k = 1, \ldots, 9$. Upper: The reproductivities $m_{s\cdot}$ and $m_{1\cdot} = \ldots = m_{s-1,\cdot} = 1/m_{s\cdot}$ of the reservoir and perimeter populations are varied, whereas $\delta = 0.2$ and $\gamma = 0$ are kept fixed. Middle: Migration from perimeter to reservoir, $0 \leq \gamma \leq 0.8$, is varied, whereas $m_{1\cdot} = 0.5$, $m_{s\cdot} = 2$ and $\delta = 0.2$ are kept fixed. Lower: Migration rate between perimeter populations, $0 \leq \delta \leq 1$, is varied, whereas $m_{1\cdot} = 0.5$, $m_{s\cdot} = 2$ and $\gamma = 0$ are kept fixed.