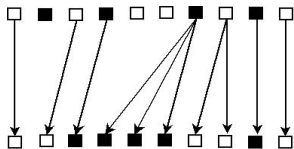# Quasi equilibrium methods in population genetics

Ola Hössjer
Dept. of Mathematics
Stockholm University

November 2012

Wright Fisher Model
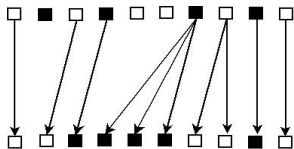
- Population of $2N$ gene (marker) copies

- Allele 1 and 2

- Nonoverlapping generations

- Constant population size

- P_t = frequency of Allele 1, Generation t,

$$
\begin{aligned}
2N &= 10, \\
P_t &= 0.4, \\
P_{t+1} &= 0.5.
\end{aligned}
$$

Wright Fisher Model

- Population of $2N$ gene (marker) copies

- Allele 1 and 2

- Nonoverlapping generations

- Constant population size

- $P_t =$ frequency of Allele 1, Generation t,

$$2N = 10,$$
$$P_t = 0.4,$$
$$P_{t+1} = 0.5.$$

Alleles drawn randomly from parental generation:

$$P_{t+1}|P_t \sim \text{Bin}(2N, P_t)/(2N)$$

Variance effective population size N_{eV}

Write

$$P_{t+1} = P_t + \varepsilon_{t+1}$$

where ε_{t+1} is genetic drift, with $E(\varepsilon_{t+1}|P_t) = 0$ and

$$\text{Var}(\varepsilon_{t+1}|P_t) = \frac{P_t(1 - P_t)}{2N_{eV}} \overset{\text{WF}}{\Longrightarrow} N_{eV} = N.$$

Variance effective population size N_{eV}

Write

$$P_{t+1} = P_t + \varepsilon_{t+1}$$

where ε_{t+1} is genetic drift, with $E(\varepsilon_{t+1}|P_t) = 0$ and

$$\text{Var}(\varepsilon_{t+1}|P_t) = \frac{P_t(1 - P_t)}{2N_{eV}} \overset{\text{WF}}{\Longrightarrow} N_{eV} = N.$$
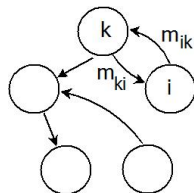
Rule of thumb:

More inbreeding $\Longrightarrow$ More genetic drift $\Longrightarrow$ Smaller N_{eV}

In general $N_{eV} \neq N$ due to

- ▶ Diploid population
- ▶ Spatial structure
- ▶ Varying reproductivity
- ▶ Time varying population size
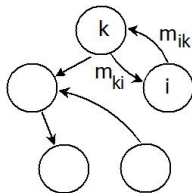- ▶ Overlapping generations

Structured population

- s subpopulations
- m_{ki} migration rate from Subpopulation k to i
- N total population size ($= 2N$ gene copies)

Structured population

- s subpopulations
- m_{ki} migration rate from Subpopulation k to i
- N total population size ($= 2N$ gene copies)

$$
\begin{aligned}
\mathbf{M} &= (m_{ki})_{k,i=1}^{s} &=& \quad \text{migration matrix} \\
Na_i &= &&\quad \text{constant subpopulation sizes}
\end{aligned}
$$

where $\sum_{i=1}^{s} a_i = 1$ and

$$
\mathbf{a} = (a_1, \dots, a_s) = \mathbf{aM},
$$

is a left eigenvector of $\mathbf{M}$ with eigenvalue 1.

Structured population

- s subpopulations
- m_{ki} migration rate from Subpopulation k to i
- N total population size ($= 2N$ gene copies)

$$
\begin{aligned}
\mathbf{M} &= (m_{ki})^s_{k,i=1} &&= \text{migration matrix} \\
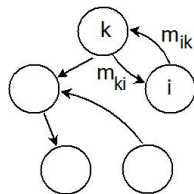Na_i &= &&\text{constant subpopulation sizes}
\end{aligned}
$$

where $\sum_{i=1}^s a_i = 1$ and

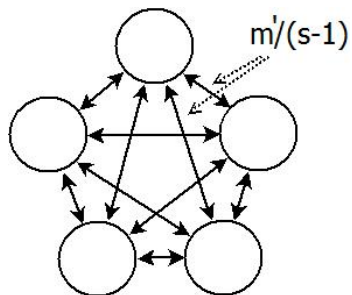$$\mathbf{a} = (a_1, \ldots, a_s) = \mathbf{aM},$$

is a left eigenvector of $\mathbf{M}$ with eigenvalue 1.

Overall migration rate $m' = \sum_{i=1}^s a_i(m_{i\cdot} - m_{ii}) = 1 - \sum_{i=1}^s a_i m_{ii}$.
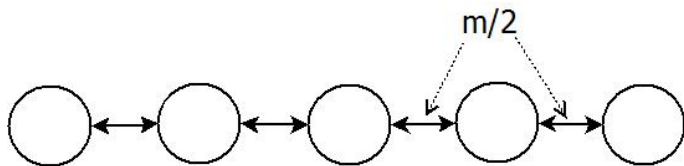
Island model

$$m_{ki} = \begin{cases} 1 - m', & i = k, \\ m'/(s-1), & i \neq k. \end{cases}$$

Linear stepping stone model

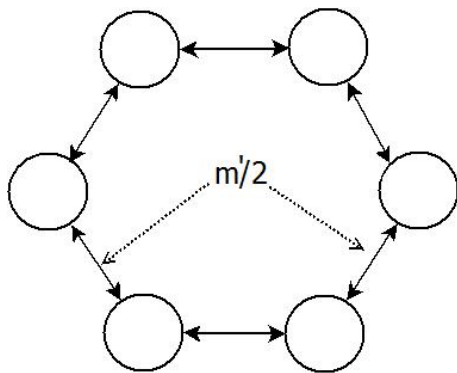$$m_{ki} = \begin{cases} 1 - m|\mathcal{N}_k|/2, & i = k, \\ m/2, & |i - k| = 1, \\ 0, & \text{otherwise}, \end{cases}$$

where $\mathcal{N}_k$ is the neighbourhood of k. Hence

$$m' = \frac{2}{s} \cdot \frac{m}{2} + \frac{s-2}{s} \cdot m = \frac{m(s-1)}{s}.$$

Circular stepping stone

$$m_{ki} = \begin{cases} 1 - m', & i = k, \\ m'/2, & (i - k \mod s) = 1 \text{ or } s - 1, \\ 0, & \text{otherwise.} \end{cases}$$

Rectangular stepping stone ($s = s_1 s_2$)

$$m_{(k_1,k_2),(i_1,i_2)} = \begin{cases} 1 - |\mathcal{N}_{(k_1,k_2)}|m/4, & (i_1,i_2) = (k_1,k_2), \\ m/4, & |i_1 - k_1| + |i_2 - k_2| = 1, \\ 0, & \text{otherwise.} \end{cases}$$

where $\mathcal{N}_{(k_1,k_2)}$ is the neighbourhood of (k_1, k_2). This yields

$$m' = m(1 - 0.5(s_1^{-1} + s_2^{-1})).$$

Torus stepping stone

$$m_{(k_1,k_2),(i_1,i_2)} = \begin{cases} 1 - m', & (i_1, i_2) = (k_1, k_2), \\ m'/4, & (i_1 - k_1 \bmod s_1) = 1 \text{ or } s_1 - 1, i_2 = k_2, \\ m'/4, & (i_2 - k_2 \bmod s_2) = 1 \text{ or } s_2 - 1, i_1 = k_1, \\ 0, & \text{otherwise.} \end{cases}$$

Backward migration matrix B

Let

$$b_{ik} = P(\text{parent of Subpop } i \text{ gene from Subpop } k) \approx \frac{a_k m_{ki}}{a_i}.$$

Backward migration matrix B

Let

$$b_{ik} = P(\text{parent of Subpop } i \text{ gene from Subpop } k) \approx \frac{a_k m_{ki}}{a_i}.$$

Since

$$\sum_{k=1}^{s} b_{ik} = \frac{1}{a_i} \sum_{k=1}^{s} a_k m_{ki} = \frac{a_i}{a_i} = 1,$$

the backward matrix

$$\mathbf{B} = (b_{ik})_{i,k=1}^{s}$$

is transition matrix of Markov chain with equilibrium distr

$$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_s).$$

Example of **M** and **B**

$$\mathbf{M} = \begin{pmatrix} 0.8 & 1.2 \\ 0.1 & 0.4 \end{pmatrix}$$

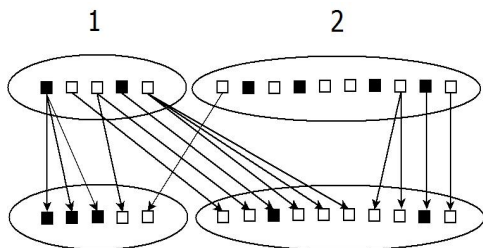$$\Downarrow$$

$$\mathbf{a} = (1/3, 2/3)$$

$$\mathbf{B} = \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix}$$

$$\Downarrow$$

$$\gamma = (3/4, 1/4).$$

It can be shown that

$$m_{k\cdot} = \sum_{i=1}^{s} m_{ki} = 1, \ k = 1, \ldots, s \Longrightarrow \gamma = \mathbf{a}.$$

Allele frequencies, fixation index

Let

$$P_{ti} = \text{frequency of Allele 1 in Subpop } i$$

and

$$P_t = \sum_{i=1}^{s} a_i P_{ti} = \text{frequency of Allele 1 in whole pop}$$

in Generation t.

Allele frequencies, fixation index

Let
$$P_{ti} = \text{frequency of Allele 1 in Subpop } i$$

and
$$P_t = \sum_{i=1}^{s} a_i P_{ti} = \text{frequency of Allele 1 in whole pop}$$

in Generation t. Then
$$F_{ST} = \text{fixation index} = \frac{\sum_{i=1}^{s} a_i (P_{ti} - P_t)^2}{P_t (1 - P_t)}$$
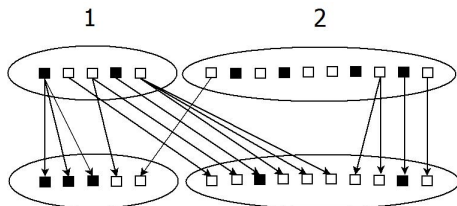
quantfies **spatial diversity of subpopulations**.

$$
\begin{array}{rcl}
\text{More migration} & \implies & \text{Smaller } F_{ST} \\
\text{More genetic drift} & \implies & \text{Larger } F_{ST}
\end{array}
$$

Allele frequency evolvement

Let

$$P_{ti} = \text{freq of Allele 1, Subpopulation } i, \text{ Generation } t$$

$$
\begin{array}{ll}
P_{t1} = 0.4 & P_{t2} = 0.4, \\
P_{t+1,1} = 0.6 & P_{t+1,2} = 0.2.
\end{array}
$$

We get the recursion

$$
\begin{aligned}
P_{t+1,i} &\approx P(\text{Parent of random gene of Subpop } i, \text{ Gen } t+1, \text{ is Allele 1}) \\
&\approx \sum_{k=1}^{s} b_{ik} P_{tk}.
\end{aligned}
$$

Allele frequency evolvement, vector form

Putting
$$\mathbf{P}_t = (P_{t1}, \ldots, P_{ts})^T,$$

we get the recursion

$$\mathbf{P}_{t+1} = \mathbf{B}\mathbf{P}_t + \varepsilon_{t+1}, \qquad (1)$$

with vector valued genetic drift term satisfying

$$E(\varepsilon_{t+1}|\mathbf{P}_t) = \mathbf{0} \text{ and } \operatorname{Var}(\varepsilon_{t+1}|\mathbf{P}_t) = \mathbf{\Sigma}(\mathbf{P}_t).$$

- (1) is vector valued heteroscedastic AR process
- Nonstationarity since $\mathbf{B}$ has largest eigenvalue 1
- $\mathbf{\Sigma}(\cdot)$ depends on reproduction scheme

Reproduction scheme 1: Fertilization precedes migration
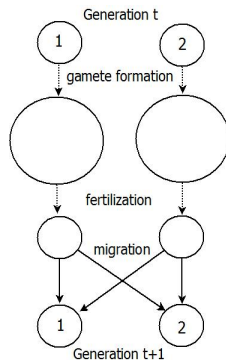
Gametes from $N_{ek} \leq Na_k$ breeders

$$\tilde{P}_{tk}|P_{tk} \sim \text{Hyp}(2Na_k, 2N_{ek}, P_{tk})/(2N_{ek}),$$

followed by **fertilization**

$$P_{tki}^*|\tilde{P}_{tk} \sim \text{Bin}(2Na_k m_{ki}, \tilde{P}_{tk})/(2Na_k m_{ki})$$

and **migration**

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

Generation t

1   2

gamete formation

fertilization

migration

1   2

Generation t+1

Reproduction scheme 2: Migration precedes fertilization

Gametes from $N_{ek} \leq Na_k$ breeders

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

followed by **migration**

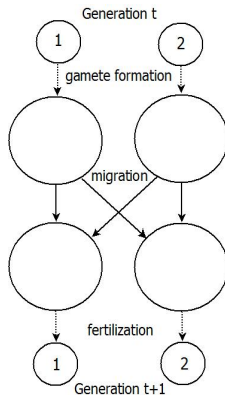$$\breve{P}_{ti} = \sum_{k=1}^{s} B_{ik} \tilde{P}_{tk},$$

where

$$(B_{i1}, \ldots, B_{is}) \sim \text{Dir}(\alpha(b_{i1}, \ldots, b_{is})),$$

and **fertilization**

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

Cointegration idea

Decompose allele frequency vector as

$$\begin{aligned}\mathbf{P}_t &= P_t^\gamma (1, \dots, 1)^T + \mathbf{P}_t^0 \\ &= \text{overall frequency } + \text{ spatial frequency fluctuations}\end{aligned}$$

where

$$P_t^\gamma = \sum_{i=1}^{s} \gamma_i P_{ti} \overset{\boldsymbol{\gamma} = \mathbf{a}}{=} P_t.$$

Cointegration idea

Decompose allele frequency vector as

$$
\begin{aligned}
\mathbf{P}_t &= P_t^{\gamma}(1,\ldots,1)^T + \mathbf{P}_t^0 \\
&= \text{overall frequency } + \text{ spatial frequency fluctuations}
\end{aligned}
$$

where

$$
P_t^{\gamma} = \sum_{i=1}^{s} \gamma_i P_{ti} \overset{\boldsymbol{\gamma}=\mathbf{a}}{=} P_t.
$$

This gives recursion

$$
\begin{cases}
P_{t+1}^{\gamma} = P_t^{\gamma} + \varepsilon_{t+1}^{\gamma}, & \text{(genetic drift part)} \\
\mathbf{P}_{t+1}^0 = \mathbf{B}^0 \mathbf{P}_t^0 + \varepsilon_{t+1}^0, & \text{(spatial fluctuation recursion part)},
\end{cases}
$$

where

$$
\varepsilon_{t+1}^{\gamma} = \sum_{i=1}^{s} \gamma_i \varepsilon_{ti}, \quad \varepsilon_{t+1}^0 = \boldsymbol{\varepsilon}_{t+1} - \varepsilon_{t+1}^{\gamma}(1,\ldots,1)^T
$$

Cointegration idea

Decompose allele frequency vector as

$$
\begin{aligned}
\mathbf{P}_t &= P_t^\gamma (1, \ldots, 1)^T + \mathbf{P}_t^0 \\
&= \text{overall frequency } + \text{ spatial frequency fluctuations}
\end{aligned}
$$

where

$$
P_t^\gamma = \sum_{i=1}^s \gamma_i P_{ti} \overset{\gamma = \mathbf{a}}{=} P_t.
$$

This gives recursion

$$
\begin{cases}
P_{t+1}^\gamma &= P_t^\gamma + \varepsilon_{t+1}^\gamma, & \text{(genetic drift part)} \\
\mathbf{P}_{t+1}^0 &= \mathbf{B}^0 \mathbf{P}_t^0 + \varepsilon_{t+1}^0, & \text{(spatial fluctuation recursion part)},
\end{cases}
$$

where

$$
\varepsilon_{t+1}^\gamma = \sum_{i=1}^s \gamma_i \varepsilon_{ti}, \quad \varepsilon_{t+1}^0 = \varepsilon_{t+1} - \varepsilon_{t+1}^\gamma (1, \ldots, 1)^T
$$

and by Perron-Frobenius' Theorem and Jordan decomposition

$$
\mathbf{B} = \mathbf{V} \begin{pmatrix} 1 & \ldots & & \\ 0 & d_2 & \ldots & \\ \vdots & & \ddots & \ldots \\ 0 & \ldots & 0 & d_s \end{pmatrix} \mathbf{V}^{-1} \text{ and } \mathbf{B}^0 = \mathbf{V} \begin{pmatrix} 0 & \ldots & & \\ 0 & d_2 & \ldots & \\ \vdots & & \ddots & \ldots \\ 0 & \ldots & 0 & d_s \end{pmatrix} \mathbf{V}^{-1}
$$

Quasi equlibrium

Dividing $\mathbf{P}_t$ by $\sqrt{P_t^{\gamma}(1 - P_t^{\gamma})}$ makes:

- ▶ Spatial fluctuation part quasi stationary.
- ▶ Genetic drift part still non-stationary.

Quasi equilibrium

Dividing $\mathbf{P}_t$ by $\sqrt{P_t^\gamma(1 - P_t^\gamma)}$ makes:

- Spatial fluctuation part quasi stationary.
- Genetic drift part still non-stationary.

The **drift covariance matrix**

$$\mathbf{\Sigma} = \frac{\text{Cov}(\mathbf{P}_{t+1}|P_t^\gamma)}{P_t^\gamma(1 - P_t^\gamma)} = \frac{E(\mathbf{\Sigma}(\mathbf{P}_t)|P_t^\gamma)}{P_t^\gamma(1 - P_t^\gamma)}$$

and **spatial fluctuation covariance matrix**

$$\mathbf{\Lambda} = \frac{\text{Cov}(\mathbf{P}_t^0|P_t^\gamma)}{P_t^\gamma(1 - P_t^\gamma)}$$

Quasi equilibrium

Dividing $\mathbf{P}_t$ by $\sqrt{P_t^{\gamma}(1 - P_t^{\gamma})}$ makes:

- Spatial fluctuation part quasi stationary.
- Genetic drift part still non-stationary.

The **drift covariance matrix**

$$\mathbf{\Sigma} = \frac{\text{Cov}(\mathbf{P}_{t+1}|P_t^{\gamma})}{P_t^{\gamma}(1 - P_t^{\gamma})} = \frac{E(\mathbf{\Sigma}(\mathbf{P}_t^0)|P_t^{\gamma})}{P_t^{\gamma}(1 - P_t^{\gamma})}$$

and **spatial fluctuation covariance matrix**

$$\mathbf{\Lambda} = \frac{\text{Cov}(\mathbf{P}_t^0|P_t^{\gamma})}{P_t^{\gamma}(1 - P_t^{\gamma})}$$

can be used to find

$$N_{eV} \approx \frac{1 - (\mathbf{a} - \gamma)\mathbf{\Lambda}(\mathbf{a} - \gamma)^T}{2\left(\mathbf{a}(\mathbf{B} - \mathbf{I})\mathbf{\Lambda}(\mathbf{B} - \mathbf{I})^T\mathbf{a}^T + \mathbf{a}\mathbf{\Sigma}\mathbf{a}^T\right)} \overset{\gamma = \mathbf{a}}{=} \frac{1}{2\mathbf{a}\mathbf{\Sigma}\mathbf{a}^T}$$

and

$$F_{ST} \approx \frac{\sum_{i=1}^{s} a_i \left((\mathbf{I} - \mathbf{1a})\mathbf{\Lambda}(\mathbf{I} - \mathbf{1a})^T\right)_{ii}}{1 - (\mathbf{a} - \gamma)\mathbf{\Lambda}(\mathbf{a} - \gamma)^T} \overset{\gamma = \mathbf{a}}{=} \sum_{i=1}^{s} a_i \Lambda_{ii}.$$

Algorithm

Linear system of equations

$$\begin{aligned}
\text{vech}(\mathbf{\Sigma}) &= \mathbf{f} - \mathbf{F}\text{vech}(\mathbf{\Lambda}), \\
\text{vech}(\mathbf{\Lambda}) &= \mathbf{G}\text{vech}(\mathbf{\Sigma}),
\end{aligned}$$

in the $s(s+1)$ unknown parameters

$$\left\{ \begin{aligned}
\text{vech}(\mathbf{\Sigma}) &= \{\Sigma_{ik}; \ i \geq k\} \\
\text{vech}(\mathbf{\Lambda}) &= \{\Lambda_{ik}; \ i \geq k\}
\end{aligned} \right.$$

with $\mathbf{F}$ and $\mathbf{G}$ square matrices of order $s(s+1)/2$ and $\mathbf{f}$ a column vector of length $s(s+1)/2$.
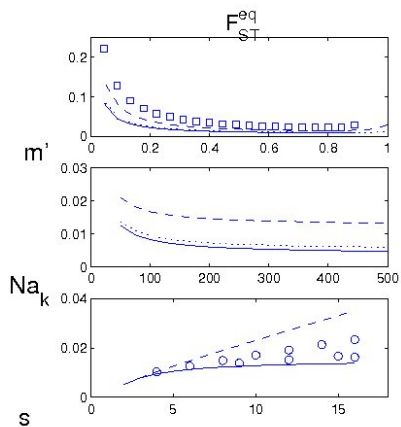
Algorithm

Linear system of equations

$$\begin{aligned}
\text{vech}(\mathbf{\Sigma}) &= \mathbf{f} - \mathbf{F}\text{vech}(\mathbf{\Lambda}), \\
\text{vech}(\mathbf{\Lambda}) &= \mathbf{G}\text{vech}(\mathbf{\Sigma}),
\end{aligned}$$

in the $s(s+1)$ unknown parameters

$$\left\{ \begin{aligned}
\text{vech}(\mathbf{\Sigma}) &= \{\Sigma_{ik}; \ i \geq k\} \\
\text{vech}(\mathbf{\Lambda}) &= \{\Lambda_{ik}; \ i \geq k\}
\end{aligned} \right.$$

with $\mathbf{F}$ and $\mathbf{G}$ square matrices of order $s(s+1)/2$ and $\mathbf{f}$ a column vector of length $s(s+1)/2$.
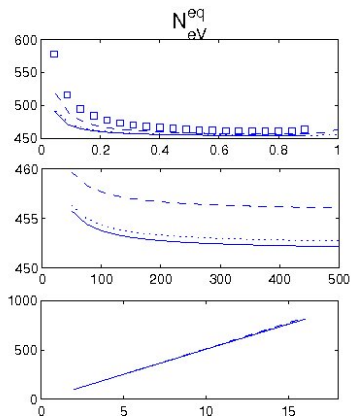
Input parameters are:

- Migration matrix $\mathbf{M} \longrightarrow \mathbf{B} \longrightarrow \mathbf{G}$
- Reproduction scenario $\longrightarrow \mathbf{\Sigma}(\cdot) \longrightarrow \mathbf{f}, \mathbf{F}$
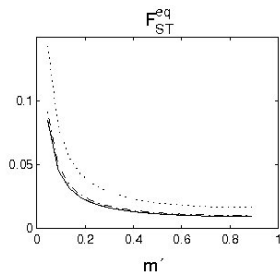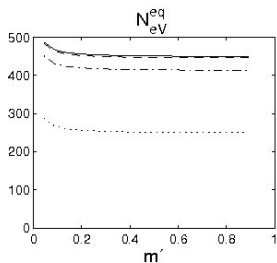
Fertilization before migration

Migration model	Symbol
Island	Solid
Torus	Dotted and circles
Circular	Dashed
Linear stepping stone	Sqaures

Fixed parameters
$N = 450$
$s = 9$
$m' = 0.4$
$a_k = 1/9$
$N_{ek} = N/9 = 50$

Migration before fertilization
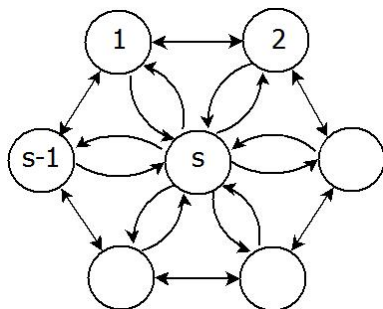
$$s = 9, N = 450, a_k = 1/9, N_{ek} = Na_k$$

Model	α	Symbol
Island	∞	Solid
	100	Dashed
	10	Dash-dotted
	1	Dotted

Model	α	Symbol
Island	10	Dash-dotted
Circular stepstone	10	Solid

Demographic reservoir (source)

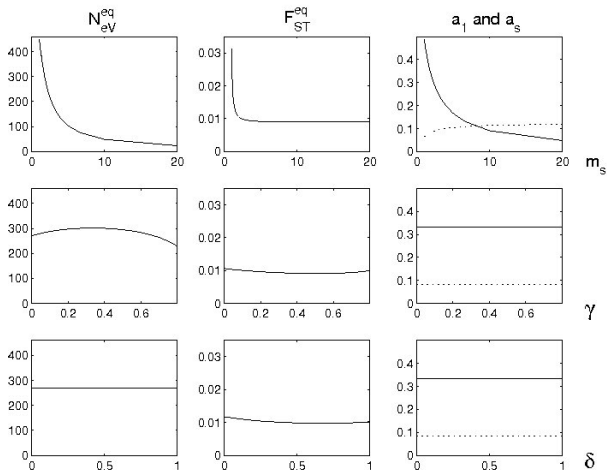$$m_{1,\cdot} = \ldots = m_{s-1,\cdot} < 1 < m_{s,\cdot}$$

and

$$m_{ki} = \begin{cases} m_s.\beta, & k = s, i < s, \\ m_s.(1 - (s-1)\beta), & k = i = s, \\ m_1.\gamma, & k < s, i = s, \\ m_1.\delta/2, & k < s, i - k = \pm 1 \bmod s - 1, \\ m_1.(1 - \gamma - \delta), & k = i < s, \\ 0, & \text{otherwise.} \end{cases}$$

Demographic reservoir, fertilization precedes migration

Top :
$m_{1\cdot} = \ldots = m_{s-1,\cdot} = 1/m_{s\cdot}$

Right :
a_1 : dotted,
a_s : solid

Fixed parameters
$s = 9$
$N = 450$
$N_{ek} = Na_k$
$\delta = 0.2$
$\gamma = 0$
$m_{1\cdot} = 0.5$
$m_{s\cdot} = 2$

Extensions

- Multipel markers (Nei's G_{ST} instead of F_{ST})
- Spatially invariant migration
    - Decompose **B** by Fourier analysis (instead of Jordan)
    - Much faster algorithm
- Spatial correlations

$$\text{Corr}(P_{ti}, P_{tk}) = \frac{\Lambda_{ik}}{\sqrt{\Lambda_{ii}\Lambda_{kk}}}$$

  at quasi equilibrium.
- Varying (sub)population sizes
- Diploid populations
- Overlapping generations

References

Hössjer, O., Jorde, P.E. and Ryman, N. (2012). Quasi equilibrium approximations of the fixation index under neutrality: The island model. To appear in *Theoretical Population Biology*.

Hössjer, O. and Ryman, N. (2012). Quasi Equilibrium, Variance Effective Population Size and Fixation Index for Models with Spatial Structure. Report 2012:4, Mathematical Statistics, Stockholm University. Submitted.

Hössjer, O. (2012). Spatial Autocorrelation for Subdivided Populations with Invariant Migration Schemes. Report 2012:11, Mathematical Statistics, Stockholm University. Submitted.

Olsson, F. Hössjer, O., Laikre, L. and Ryman, N. (2012). On the Variance Effective Size of Fluctuating Populations with Overlapping Generations. In preparation.