Genetic Exploration of Substructured Populations
GESP: A User's Manual

Version 1.0.1

Fredrik Olsson
GESP.info@gmail.com

February 26, 2017

Citation:

Olsson F, Laikre L, Hössjer O, Ryman N (2016) GESP: A computer program for modeling genetic effective population size, inbreeding, and divergence in substructured populations. Under revision for Molecular Ecology Resources.

Description:

GESP (Genetic Exploration of Substructured Populations) is an R-based software designed to model short and long term patterns of genetic differentiation and effective population size of subdivided populations. It applies newly developed analytical approaches that allow varying number and size of subpopulations, migration schemes, and degree of genetic relatedness among and within populations. Output quantities all have analytical closed form expressions and include for example local and global inbreeding effective population size that can be measured over different time intervals, eigenvalue effective size, g_{ST}, inbreeding and kinship coefficients. The mathematical theory behind GESP is provided in Hössjer et al. (2014, 2015) and Olsson et al. (2016). GESP has been developed as a tool for conservation geneticists as exemplified in Olsson et al. (2016) and Laikre et al. (2016).

Contents

1   Getting started

In this section we describe how to obtain, install and start GESP. Usually, administrative rights are needed in order to install the programs.

1.1   Prerequisites

The program GESP depends on a number of other programs which must be installed prior to the installation of GESP:

R Visit cran.r-project.org to download and install the latest version of R.

GTK+ The graphical user interface depends on GTK+ (www.gtk.org). For Linux users, GTK+ is usually installed by default (e.g. Ubuntu and Debian). If you are running another operating system, GTK+ needs to be installed separately. On the operation systems

- Windows 7
- Windows 10 pro - x64
- Mac OS 10.7.5

the following installation process has been successfully tested

- Start R.
- Write the following commands at the command line prompt:

  ```
  install.packages("RGtk2",depen=T)
  library("RGtk2")
  ```

- If GTK+ is not already installed, R will show a pop-up dialog asking you to install it. Accept and follow the instructions on the screen.
- When GTK+ is installed, restart R.
- To confirm that GTK+ is installed, no error message should be shown when the following command is written at the command line prompt:

  ```
  library("RGtk2")
  ```

  For other operation systems or if you encounter any problems while installing GTK+, instructions at https://gist.github.com/sebkopf/9405675 and http://www.learnanalytics.in/blog/?p=31 are very helpful.

When both R and GTK+ are installed, a number of R-packages (gWidgets, gWidgetsRGtk2, RGtk2, cairoDevice) need to be installed in order for GESP to function. These packages can be installed in various ways, but it is easiest to install them from within R in the following way:

- Start R.

- Write the following command at the command line prompt:

  ```
  install.packages(c("gWidgets","gWidgetsRGtk2","RGtk2","cairoDevice"))
  ```

- In the pop-up dialog, choose an appropriate mirror (one that is close to you).

1.2 Installation

GESP can be downloaded from www.zoologi.su.se/research/GESP/. To install GESP, start R and write the following command at the command line prompt:

```
install.packages("PATH-TO-FILE",repos=NULL,type="sources")
```

where "PATH-TO-FILE" should be replaced by the path to the GESP source file. Example: For a windows user "Fredrik" who put GESP version 1.0 on the desktop, using the following command GESP will be installed:

```
install.packages("C:/Documents and
Settings/Fredrik/Desktop/GESP_1.0.tar.gz",repos=NULL,type="sources")
```

During the installation of GESP and the prerequisites, R might show warning or error messages. If GESP does not start as after the installation process, restarting R might resolve the issues.

1.3 Starting the program

In order to run the program, first make sure that GESP and all prerequisites are successfully installed. Then, at the R command line write the following:

```
library(GESP)
startGESP()
```

This will bring up the main interface.

1.4 The Interface

Figure 1 shows the main interface of the program. It consists of the following parts:

File (1) In this menu a new model can be initiated, loaded and saved, and it is also possible to quit the application. For windows users there is an option to copy the result table to the clipboard see Section 2 and 6.

Run (2) In this menu, the specified model can be processed in order to obtain results see Section 3.

Model parameters (3) In this tab the model parameters are specified, see Section 2.2.

Result options (4) This tab contains various result options that can be chosen for a specified model, see Section 3.

Information and result tabs (5) Here, various results and information about the specified model are presented, see Section 5.

Status bar (6) The status bar contains information about the latest command and the current model.
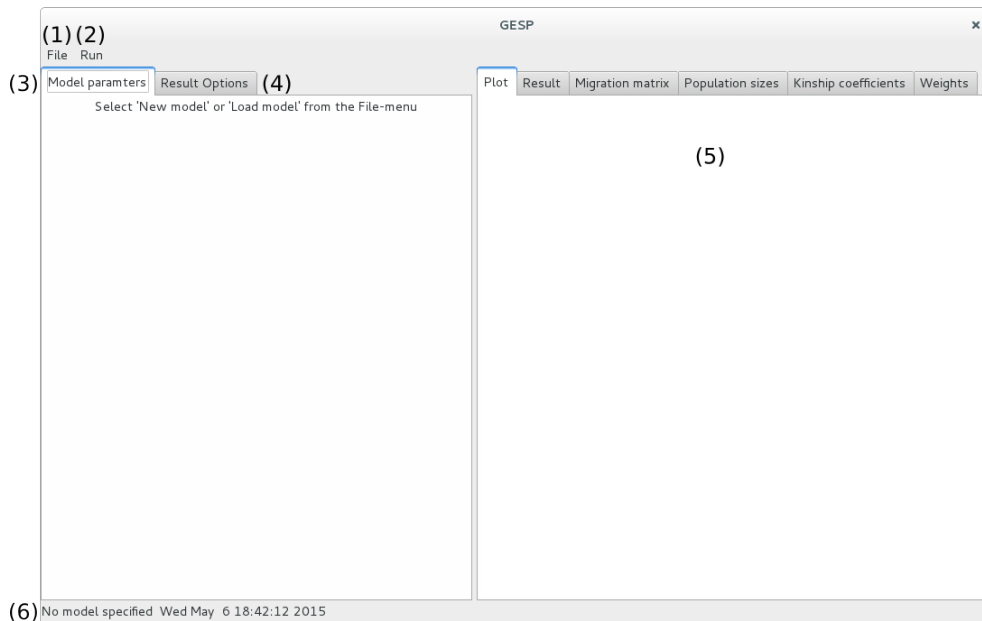
Figure 1: The main interface.

2 Specifying a new model

This section provides a step by step instruction on how to specify a new model and the various options in the interface. For a detailed description on how to specify a migration scheme, please go to Section 4. In Section 8 a number of examples are presented which can be loaded into GESP.

In order to specify a new model, first choose "New model" from the "File"-menu. This will bring up the New model window, shown in Figure 2. In this window the number of subpopulations, ploidy and migration type is specified which can only be altered by specifying a new model. Other choices that are made in this window are present as a help for the user to specify the initial model and can be altered at a later stage in the program.

2.1 New model

Number of subpopulations Specifies the number of subpopulations that should be included in the model. This parameter is called s in Hössjer et al. (2014, 2015), and can assume values between 2 and 100.

Initial census sizes Specifies the initial local census sizes of all subpopulations. This parameter is called N_i (without time index) in Hössjer et al. (2014, 2015), and can assume values between 1 and 10^8. This parameter can be overridden at a later stage of the program, making it possible vary the local census size N_{ti} over time t.

Initial effective sizes Specifies the initial local effective sizes of all subpopulations. This parameter is called N_{ei} (without time index) in Hössjer et al.

Figure 2: The new model window.

(2014, 2015), and can assume values between 1 and 10^8. This parameter can be modified at a later stage of the program, making it possible vary the local effective size N_{eti} over time t, in such a way that N_{eti}/N_{ti} is held constant. .

Initial kinship within subpops Specifies the initial kinship coefficient of all subpopulations. This parameter is called f_{ii} (without time index) in Hössjer et al. (2014, 2015), and can assume values between 0 and 1. This parameter can be modified at a later stage in the program.

Initial kinship between subpops Specifies the initial kinship coefficient between all subpopulations. This parameter is called f_{ij} (without time index) in Hössjer et al. (2014, 2015), and can assume values between 0 and 1. This parameter can be modified at a later stage in the program.

Migration model Specifies the migration model of the population. Predefined models include the island model, the circular stepping model and the linear stepping stone model. The user defined model of the current version is an island model.

6

Migration rate Specifies an initial migration rate. This parameter is called m in Hössjer et al. (2014), and can assume values between 0 and 1. The migration scheme, including the migration rate, are the initial settings and can be overridden at a later stage in the program.

Migration type Specifies if migration is fixed or stochastic. For fixed migration, the migration rates define the exact number of migrants, whereas for stochastic migration, the migration rates define the average number of migrants.

Ploidy Specifies whether individuals are haploid or diploid.

2.2 Parameters

In the left part of the interface there will be a parameters tab as soon a new model is loaded. In this tab, further specifications of the model can be made and the appearance differs slightly between the haploid and diploid models.

2.2.1 Haploid

When a haploid model is chosen, the interface will appear as in Figure 3. In the parameters tab, the following settings can be adjusted:
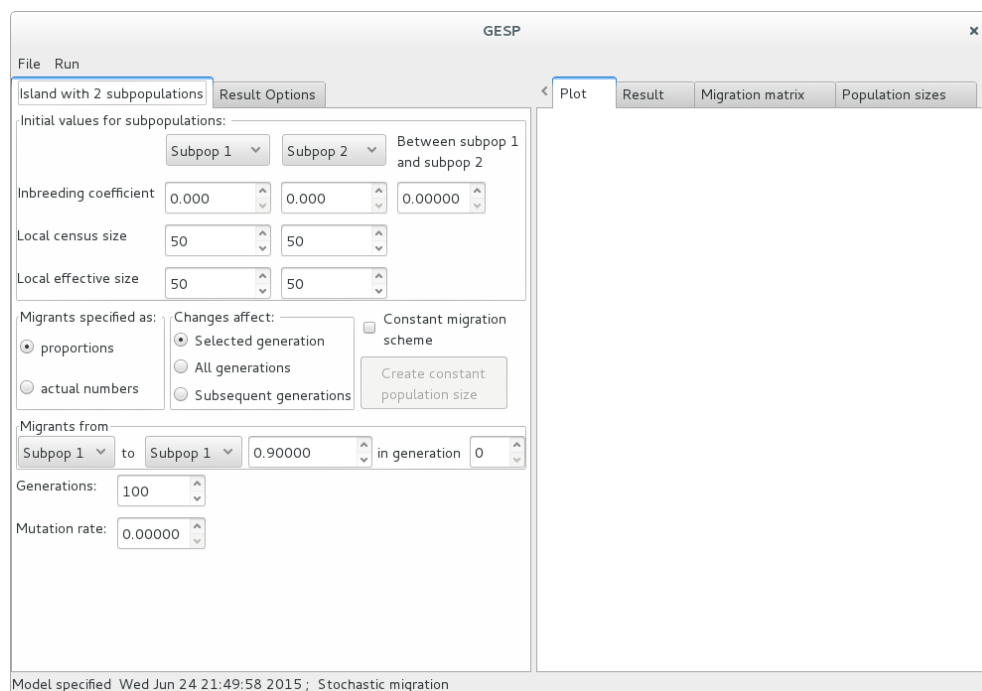

Figure 3: The main interface of a model with haploid individuals.

Inbreeding coefficient The first row of boxes specify the probability that two genes, drawn at random without replacement from within population 1

and 2, respectively, are identical (first boxes below Subpop 1 and Subpop 2; Figure 3). The third box on the first row specifies the probability that two genes drawn at random from each of subpopulation 1 and 2 are identical.

Local census size The two boxes specify the local sizes of the subpopulations in the drop-down boxes above.

Local effective size The two boxes specify the local effective sizes of the subpopulations in the drop-down boxes above. It is the effective size that the subpopulation would have if it was isolated.

Migrants specified as There are two ways to specify migration between subpopulations, by proportions or by number of individuals, see Section 4.

Changes affect This option specifies which generations that are affected by changes to the migration scheme, see Section 4.

Constant migration scheme By checking this box, the migration scheme between generation 0 and 1 are copied to all subsequent generations, see Section 4.

Create constant population size By clicking this button, the program tries to adjust all the local census sizes to a constant, see Section 4. This button is enabled by checking the "Constant migration scheme"-box.

Migrants from Here, the migration scheme is specified, see Section 4.

Generations Specify the number of generations $t_{max}+1$. Possible values range from 2 to 1000. Generations are numbered as $t = 0, 1, \ldots, t_{max}$.

Mutation rate Specify the mutation rate. This parameter is called μ in Hössjer et al. (2014), and can assume values between 0 and 1.

2.2.2 Diploid

The parameters tab for a diploid model, shown in Figure 4, differs from the parameters tab for a haploid model in the following way:

Kinship coefficient The probability f_{ij} that two genes are identical by descent IBD (or identical by state, IBS) when drawn randomly from different individuals of subpopulations i and j.

Inbreeding coefficient The probability f_i that two genes are identical by descent (or state) when drawn randomly from the same individual of subpopulation i. When a new model is specified, the default value is equal to the kinship coefficient.

Selfing The probability that an individual mates with itself. Default values are zero (no selfing) or random selfing (inverse subpopulation size).

Mating When two individuals produce offspring, they either mate before or after migration. In the former case, the pair belongs to the same subpopulation, but their offspring may migrate to other subpopulations. In the latter case, the migration phase precedes mating, and a breeding pair may originate from different subpopulations.
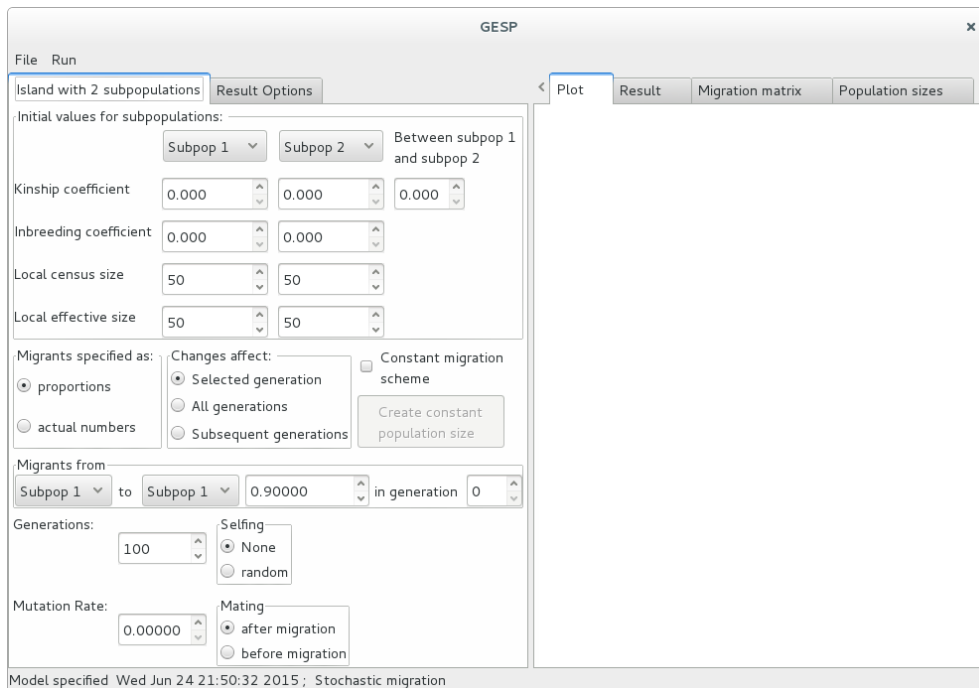
Figure 4: The main interface of a model with diploid individuals.

3 Results

When the model has been specified, press tab to ensure that GESP registers the last edit of the model, then choose "Run analysis" from the run menu in order to obtain results.

Choose result type The results can either be presented using effective population sizes or in terms of inbreeding and kinship coefficients and subpopulation differentiation.

Weighting scheme A weight w_i is assigned to each subpopulation i in terms of the probability of sampling a single gene from it. Pre-specified weighting schemes are uniform weights ($w_i = 1/s$) i.e. equal or weights proportional to local census sizes.

User defined weights This option makes it possible to specify arbitrary subpopulation weights, including local weights, where one subpopulation i has full weight ($w_i = 1$).

Choosing result parameters Choose which result parameters that should be shown. The various choices are explained below.

NeE The eigenvalue effective size N_{eE}. It quantifies the long term rate at which inbreeding increases in the whole population. This box is enabled only if the "Constant migration scheme"-box is checked and all subpopulation sizes are constant.
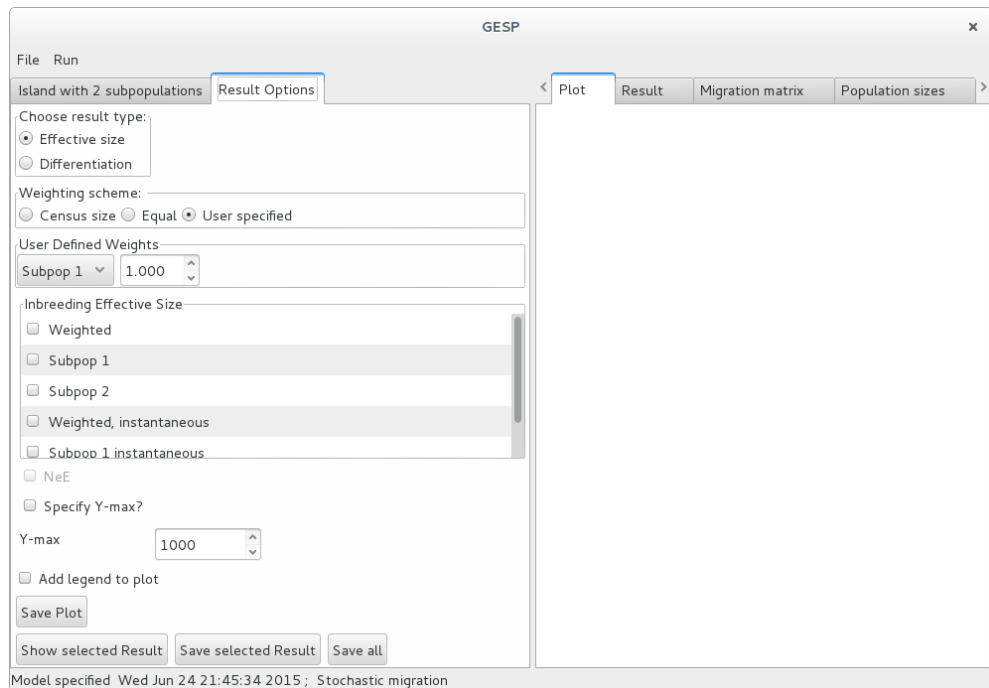
9

Figure 5: The result option tab.

Specify Y-max Check this box if the maximum value of the y-axis should be user specified.

Y-max The maximum value of the y-axis if the "Specify Y-max" checkbox is checked.

Add legend to plot If this box is ticked, a legend is added when the result are plotted.

Save plot Press this button in order to save the current plot. The format of the figure is determined by the file extension and the possible formats are

- pdf
- png
- ps
- jpg
- tiff

For example, naming the output file 'results.pdf' will produce a pdf file with the current plot.

Show selected result Press this button to show the selected result in the Result tab.

Save selected result Press this button to save the selected result in a csv-file.

Save all Press this button to save all possible results in two csv-files.

3.1 Haploid

3.1.1 Effective size

The following effective size curves employ inbreeding probabilities for pairs of genes drawn randomly without replacement from the total population. The randomness mechanism is determined by the probabilities of drawing genes from different subpopulations (the subpopulation weights).

Weighted A curve for the inbreeding effective size N_{eI} over intervals $[0, t]$ starting at 0 when subpopulations are weighted uniformly, proportional to local census size or in a user-specified way ($t = 1, 2, \ldots, t_{max}$).

Subpop i A curve for the local inbreeding effective sizes N_{eI} of subpopulation i, over intervals $[0, t]$ starting at 0 ($t = 1, 2, \ldots, t_{max}$).

Weighted, instantaneous A curve for the instantaneous inbreeding effective sizes N_{eI} between consecutive generations $[t, t+1]$, when subpopulations are weighted uniformly, proportional to local census size or in a user-specified way ($t = 0, 1, \ldots, t_{max} - 1$).

Subpop i instantaneous A curve for the local inbreeding effective sizes N_{eI} of subpopulation i between consecutive generations $[t, t+1]$ ($t = 0, 1, \ldots, t_{max} - 1$).

3.1.2 Differentiation and kinship

The quantities below are all computed for $t_{max}+1$ generations ($t = 0, 1, \ldots, t_{max}$). They represent theoretically expected (or predicted) values of each quantity. For ease of notation, subscript t is often omitted. The kinship coefficients and coefficients of gene differentiation involve drawing a gene pair randomly from the population, with our without replacement, as specified below. The randomness mechanism is determined by the probabilities of drawing genes from different subpopulations (the subpopulation weights).

gST Prediction $g_{ST} = g_{STt} = (h_{Tt} - h_{St})/h_{Tt}$ of the coefficient of gene differentiation $G_{ST} = G_{STt}$. It is based on drawing genes with replacement, as defined by Nei (1973) for uniform weights, and Nei (1977) for general weighting schemes.

gST' A normalized version of g_{ST} defined by Hedrick (2005) and Meirmans and Hedrick (2011). Its upper bound of 1 can be attained for any choice of kinship coefficients within subpopulations. It is based on drawing genes with replacement.

fS Kinship coefficient $f_S = f_{St}$ for a randomly chosen gene pair from the same subpopulation (S). It is a weighted average of the kinship coefficients f_{ii} of all subpopulations i. It is based on drawing genes with replacement.

fT Kinship coefficient $f_T = f_{Tt}$ for a randomly chosen gene pair from the total population (T). It is a weighted average of the kinship coefficients f_{ii} of all subpopulations i and the kinship coefficients f_{ij} for all pairs i, j of different subpopulations. It is based on drawing genes with replacement.

hS Heterozygosity $h_S = h_{St} = 1 - f_S$ for a randomly chosen gene pair from the same subpopulation (S). It is a weighted average of the heterozygosities h_{ii} of all subpopulations i. It is based on drawing genes with replacement.

hT Gene diversity $h_T = h_{Tt} = 1 - f_T$ for a randomly chosen gene pair from the total population (T). It is a weighted average of the heterozygosities h_{ii} of all subpopulations i and the gene diversities h_{ij} of all pairs i, j of different subpopulations. It is based on drawing genes with replacement.

fi,i Kinship coefficient $f_{ii} = f_{tii}$ for a randomly chosen gene pair from the same subpopulation i. It is based on drawing genes without replacement.

fi,j Kinship coefficient $f_{ij} = f_{tij}$ for a randomly chosen gene pair from different subpopulations $i \neq j$. It is based on drawing genes without replacement.

hi,i Heterozygosity, i.e. non-IBD probability $h_{ii} = h_{tii} = 1 - f_{ii}$ for a randomly chosen gene pair from the same subpopulation i. It is based on drawing genes without replacement.

hi,j Gene diversity, i.e. non-IBD probability $h_{ij} = h_{tij} = 1 - f_{ij}$ for a randomly chosen gene pair from different subpopulations $i \neq j$. It is based on drawing genes without replacement.

3.2 Diploid

3.2.1 Effective size

The following effective size curves employ inbreeding probabilities for pairs of genes drawn without replacement from the same randomly chosen individual of the population. The randomness mechanism is determined by the probabilities of drawing genes from different subpopulations (the subpopulation weights).

Weighted A curve for the inbreeding effective size N_{eI} over intervals $[0, t]$ starting at 0 when subpopulations are weighted uniformly, proportional to size or in a user-specified way ($t = 1, 2, \ldots, t_{max}$).

Subpop i A curve for the local inbreeding effective sizes N_{eI} of subpopulation i, over intervals $[0, t]$ starting at 0 ($t = 1, 2, \ldots, t_{max}$).

Weighted, instantaneous A curve for the instantaneous inbreeding effective sizes N_{eI} between consecutive generations $[t, t + 1]$, when subpopulations are weighted uniformly, proportional to size or in a user-specified way ($t = 0, 1, \ldots, t_{max} - 1$).

Subpop i instantaneous A curve for the local inbreeding effective size N_{eI} of subpopulation i between consecutive generations $[t, t + 1]$ ($t = 0, 1, \ldots, t_{max} - 1$).

3.2.2 Differentiation, kinship and inbreeding

The quantities below are all computed for $t_{max} + 1$ generations ($t = 0, 1, \ldots, t_{max}$). They represent theoretically expected (or predicted) values of each quantity. For ease of notation, subscript t is often omitted. The inbreeding coefficients,

kinship coefficients and non-IBD probabilities involve drawing a gene pair randomly from the population, with or without replacement, as specified below. The randomness mechanism is determined by the probabilities of drawing genes from different subpopulations (the subpopulation weights).

gST Prediction $g_{ST} = g_{STt} = (h_{Tt} - h_{St})/h_{Tt}$ of the coefficient of gene differentiation $G_{ST} = G_{STt}$. It is based on drawing genes with replacement, as defined by Nei (1973) for uniform weights, and Nei (1977) for general weighting schemes. .

gST' A normalized version of g_{ST} defined by Hedrick (2005) and Meirmans and Hedrick (2011). Its upper bound of 1 can be attained for any choice of kinship coefficients within subpopulations. It is based on drawing genes with replacement.

fIS Prediction $f_{IS} = f_{ISt}$ of the fixation index $F_{IS} = F_{ISt}$ of individuals (I) within subpopulations (S). It is based on drawing genes with replacement. A value of 0 signifies Hardy-Weinberg genotype proportions within subpopulations, whereas a negative (positive) value signifies an increased (decreased) number of heterozygous genotypes compared to HW proportions.

fIT Prediction $f_{IT} = f_{ITt}$ of the fixation index $F_{IT} = F_{ITt}$ of individuals (I) within the total population (T). It is based on drawing genes with replacement, and it is close to g_{ST} when genotype frequencies within subpopulations are close to Hardy Weinberg proportions.

fI Inbreeding coefficient $f_I = f_{It}$ for a gene pair of a randomly chosen individual (I) in the population. It is a weighted average of the inbreeding coefficients f_i of all subpopulations i. It is based on drawing genes with replacement.

fS Kinship coefficient $f_S = f_{St}$ for a randomly chosen gene pair in the same subpopulation (S). It is a weighted average of the inbreeding coefficients f_i and kinship coefficients f_{ii} of all subpopulations i. It is based on drawing genes with replacement.

fT Kinship coefficient $f_T = f_{Tt}$ for a randomly chosen gene pair of the total population (T). It is a weighted average of the inbreeding coefficients f_i and kinship coefficients f_{ii} of all subpopulations i, as well as the kinship coefficients f_{ij} of all pairs i, j of different subpopulations. It is based on drawing genes with replacement.

hI Heterozygosity $h_I = h_{It} = 1 - f_I$ for a gene pair of a randomly chosen individual (I) in the population. It is a weighted average of the heterozygosities h_i of all subpopulations i. It is based on drawing genes with replacement.

hS Non-IBD probability $h_S = h_{St} = 1 - f_S$ for a randomly chosen gene pair from the same subpopulation (S). It is a weighted average of the heterozygosities h_i and non-IBD probabilities h_{ii} of all subpopulations i. It is based on drawing genes with replacement.

hT Non-IBD probability $h_T = h_{Tt} = 1 - f_T$ for a randomly chosen gene pair from the total population (T). It is a weighted average of the heterozygosities h_i and non-IBD probabilities h_{ii} of all subpopulations i, as well as the non-IBD probabilities h_{ij} of all pairs i, j of different subpopulations. It is based on drawing genes with replacement.

fi Inbreeding coefficient $f_i = f_{ti}$ for a gene pair from a randomly chosen individual of subpopulation i. It is based on drawing genes without replacement.

fi,i Kinship coefficient $f_{ii} = f_{tii}$ for a randomly chosen gene pair from different individuals of the same subpopulation i. It is based on drawing genes without replacement.

fi,j Kinship coefficient $f_{ij} = f_{tij}$ for a randomly chosen gene pair from different subpopulations $i \neq j$. It is based on drawing genes without replacement.

hi Heterozygosity, i.e. a non-IBD probability $h_i = h_{ti} = 1 - f_i$ for a gene pair from a randomly chosen individual of subpopulation i. It is based on drawing genes without replacement.

hi,i Non-IBD probability $h_{ii} = h_{tii} = 1 - f_{ii}$ for a randomly chosen gene pair from different individuals of the same subpopulation i. It is based on drawing genes without replacement.

hi,j Non-IBD probability $h_{ij} = h_{tij} = 1 - f_{ij}$ for a randomly chosen gene pair from different subpopulations $i \neq j$. It is based on drawing genes without replacement.

4 Specify Migration

Migration can be specified in various ways, as described in this section and illustrated in Section 8 for some examples.

4.1 Migrants specified as

There are two ways of specifying migration, either as proportions or actual numbers. If migrants are specified as proportions, the migration rates describes the proportion of the individuals in the donor subpopulation that migrate to other subpopulations including itself. If migrants are specified as actual numbers; the migration rates describes the number of individuals that migrate from the donor population to other subpopulations (including itself).

4.2 Changes affect

There are three options, either the changes to migration scheme affect the "Specified generation" or "All generations". If the "Subsequent generations" option is chosen, changes to the migration scheme affect the current and all future generations.

4.3 Constant migration scheme

Check this box if the migration matrix should be constant over time. All changes to the migration affects all generations in the same way.

4.4 Create constant population sizes

This routine is enabled if the "Constant migration scheme" box is checked. By checking this box, the program modifies the migration matrix in order to create constant population sizes. Pressing this button apply changes to the migration matrix immediately which cannot be undone by other means than editing the migration matrix manually.

4.5 Migrants from

It is in this box that the migration matrices are specified. In the first drop-down box the donor subpopulation is chosen and in the second drop-down box the receiver subpopulation is chosen. In the next box the migration rate is chosen and the value depends on whether migrants are specified as proportions or actual numbers. The following box, "In generation", specifies which generation that is currently used.

For example, if we specify migration as actual numbers, let changes affect subsequent generations and input a migration of 40 from subpopulation 1 to subpopulation 2 in generation 5. Then, all migration matrices from generation 5 and onward will change the migration to 40 individuals between subpopulation 1 and 2.

4.6 Information tabs

On the right hand side of the interface there are two information tabs that might be of interest when migration is specified. The migration matrix tab shows the migration matrix in the specified generation. The population sizes tab shows the population sizes of all subpopulations at all generations. These tabs are automatically updated when changes to the migration matrix are made.

5 Plots, results and various information tabs

Plot In this tab the plots are shown when the Show plot button is pressed.

Result This tab shows a table of the results when the Show selected results button is pressed.

Migration matrix This tab shows the migration matrix for the current generation, see Section 4.

Population sizes This tab shows the subpopulation sizes.

Kinship coefficients In this tab the kinship coefficients are presented in a matrix.

Weights This tab contains a table of the weights used for calculating various results, see Section 3.

6 Load and save

It is possible to both load and save models specifications and it is done from the File-menu.

In order to save a model, choose "Save model" from the File-menu and enter an appropriate name. This will create a file in which all specifications done in the parameters tab are stored and with the file extention ".GESPmodel".

To load a model, choose "Load model" from the File-menu and choose the .GESPmodel file that you want to load. It will be shown in the status bar whether the load/save was successful.

7 Runtime performance

The time consumption for the "Run Analysis" command is mainly affected by the number of subpopulations and the number of generations of the model. In Table 1, the execution time for the "Run analysis" command is presented for a diploid model with different choices of the number of subpopoulations and the number of generations. In the tests, a laptop with an Intel(R) Core(TM) i7-3517U CPU @ 1.90GHz and 10 GB RAM running Debian GNU/Linux 8 was used.

Table 1: Runtime for a diploid model with different number of subpopulations and generations.

Subpopulations	Generations	Execution time
10	100	0.2 s
10	1000	1.2 s
30	100	12 s
30	1000	57 s
50	100	39 s
50	1000	340 s
75	100	170 s
75	1000	1700 s

8 Examples

In this section we present some examples to illustrate how the program can be used. We describe how to specify the model step by step. For each example a .GESPmodel file with the corresponding model specifications is available for downloading on the program homepage (www.zoologi.su.se/research/GESP/).

8.1 Island model

We start by defining an island model. Choose "New model" from the file menu. Change the number of subpops to 5, migration rate to 0.05, census size to 100, local effective sizes to 100 and ploidy to diploid. By pressing the "OK" button, the new model is created. This model can be loaded by choosing "Load model"

from the file menu and selecting the Islandmodel.GESPmodel file, which can be obtained from the program's webpage.

The parameters tab will now be called "Island with 5 subpopulations". Check the "Constant migration scheme" box. This will enable calculation of the eigenvalue effective population size if the subpopulation sizes are constant (which they are in our example). Now, choose "Run analysis" from the run menu and press the result options tab.

Check the boxes "Weighted", "Weighted, instantaneous" and "NeE", and press the "Show selected result" button. Figure 6 will then be shown to the right in the plot tab. Since we have an island model, the rate at which inbreeding increases will be the same in all subpopulations. Hence, the inbreeding effective population size for subpopulation 1 (which is shown if check-box "subpop 1" is selected) equals the weighted inbreeding effective population size.
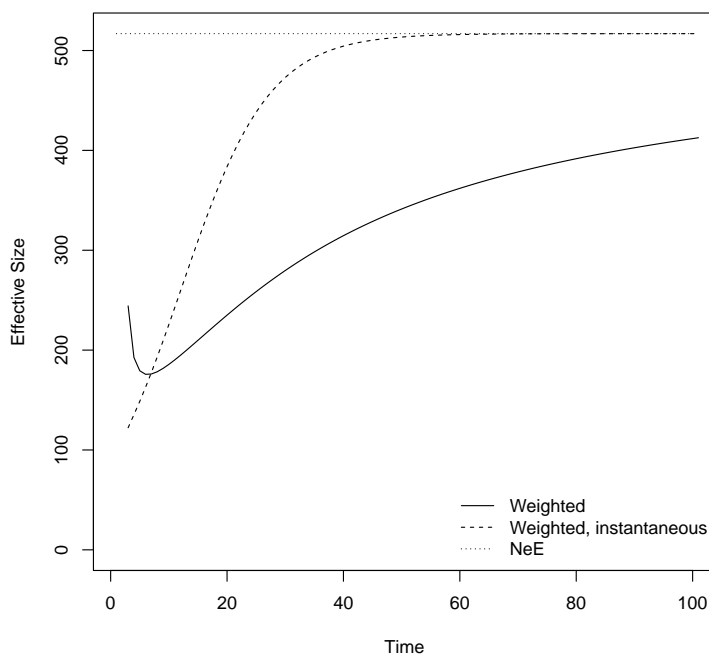

Figure 6: Local inbreeding effective sizes, and eigenvalue effective size of the metapopulation, for the island model example in Section 8.1.

In order to plot predicted kinship and inbreeding coefficients, choose "Differentiation" from the choose result type option box. Check the "gST", "f1,1", "f1,2" and "f1" boxes, and press show selected result. Figure 7 will then be shown in the plot tab to the right.
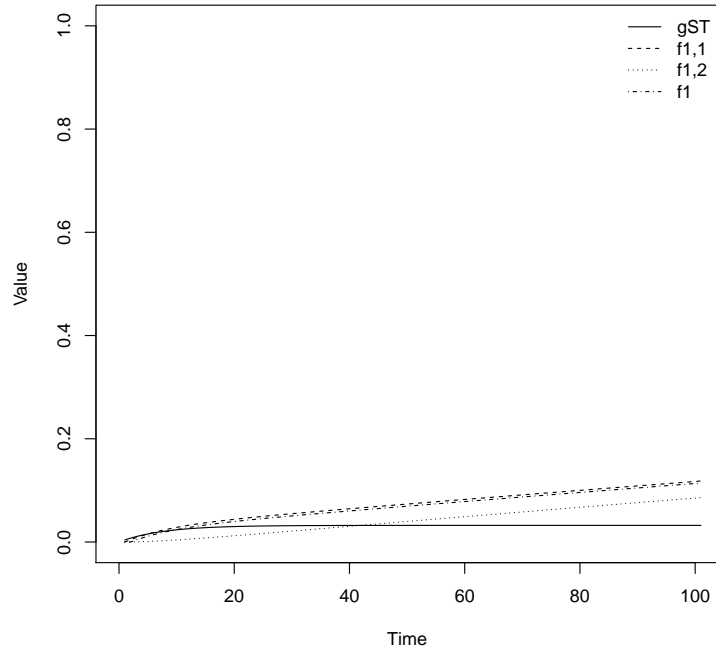
17

Figure 7: Kinship coefficients, inbreeding coefficient and g_{ST} in the island model example in Section 8.1.

8.2 Source-sink system

We will now continue by altering the model of the previous example. Start with this model and make sure that you are in the parameters tab called "Island with 5 subpopulations". We will alter the migration scheme to create a source-sink like population model. On the right side of the interface, choose the tab migration matrix. Here, the migration matrix is shown. Its element on row i and column j corresponds to the proportion of individuals living in subpopulation i at time 0 that will migrate to subpopulation j the next generation. Make sure that the "Constant migration scheme" check-box is marked, so that the migration matrix is constant over time.

In the "Migrants from" option box, change the to-box, i.e. the second box, which currently reads subpop 1, to subpop 2 and change the value 0.01250 to 0. This mean that no migration will occur from subpopulation 1 to subpopulation 2. Continue by changing the to box to subpop 3,4 and 5, and change the migration rate to 0. If you view the migration matrix on the right hand side of the interface, the first row should start by 0.95 followed by four zeros. Hence, no individual will migrate from subpopulation 1.

Switch from the migration matrix tab to the population sizes tab. Here you see the subpopulation sizes over time. Since migration from subpopulation 1

has been removed, we see that the subpopulation sizes will decrease over time. In order to create constant subpopulation sizes, switch back to the migration matrix tab.

In the "Migration from" option box, set both the from-box and to-box to sub-pop 2 and change the migration rate to 0.9625. Repeat the previous step for subpopulation 3,4 and 5. Then, all elements in the diagonal of the migration matrix should be 0.9625. Switch back to the population sizes tab. The subpopulation sizes should now be constant (100) for all subpopulations. By symmetry, subpopulations 2-5 will have the same rate of increased inbreeding, while subpopulation 1 will have another rate, since this subpopulation only receives immigrants.

The model can be loaded by choosing "Load model" from the file menu and selecting the file sourcesink.GESPmodel. In order to examine how this change of migration scheme will affect the subpopulations, choose "Run analysis" from the run menu and press the result options tab.

Check the boxes "Subpop 1" and "Subpop 2", and press the "Show selected result" button. Figure 8 will then be shown in the plot tab, to the right. We see that the inbreeding effective population size of subpopulation 1 is greater than the inbreeding effective population size of subpopulation 2.

Choose "Differentiation" from the "Choose result option" box and check the boxes "f1,1", "f1,2" and "f2,2". When the show selected result button is pressed, Figure 9 will appear in the plot tab. This figure shows the kinship coefficients of subpopulation 1 and subpopulation 2 as well as the kinship coefficient between these two subpopulations. If instead the "f1" and "f2" boxes are checked, we see that the inbreeding coefficient of subpopulation 1 will be lower than the inbreeding coefficient of subpopulation 2.

8.3   Complicated migration patterns

In this section we will analyze the more complicated migration system shown in Figure 10. The model can also be found from the complicated.GESPmodel file of the program's homepage.

Start with the model of the last example. In the "Initial values for subpopulations" section in the parameters tab, the initial values for the subpopulations are chosen. Select the subpopulations in the top row and set the values according to Table 2. The value in the first (second) column corresponds to the subpopulation in the first (second) option box and the kinship coefficient in the third column corresponds to the kinship coefficient between the subpopulations in the option boxes. To specify the migration scheme, set the "Migration specified as" to "Actual numbers" in order to define migration in terms of number of individuals that migrate. Then, set the migration scheme according to Table 3 and make sure the the "Constant migration scheme" check-box is marked. In order to check that the migration is specified correctly, switch to the migration matrix tab on the right hand side of the interface. When the migration matrix is specified, switch to the "Population sizes" tab to check that the subpopulation sizes are constant. Now choose "Run analysis" from the run menu and switch to the "Result options" tab.
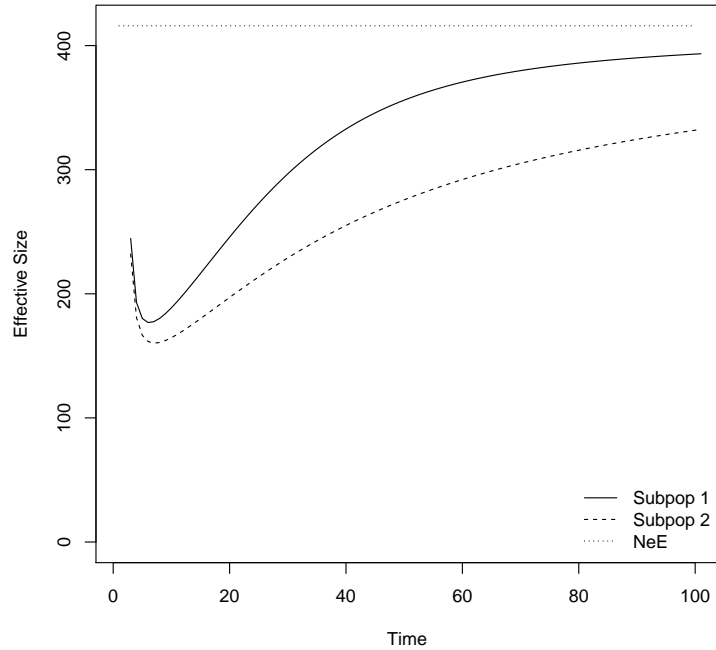
Figure 8: Inbreeding effective population sizes of subpopulations 1 and 2, and eigenvalue effective populaton size of the metapopulation, for the source sink example.

Table 2: Initial values of the parameters for all subpopulations of the complicated migration patterns example.

Subpopulation	1	2	3	4	5
Kinship coefficient	0.05	0.1	0	0.05	0
Inbreeding coefficient	0.05	0.1	0	0.05	0
Local census size	50	100	200	400	200
Local effective size	50	100	200	400	200

Choose "Differentiation" from the "Choose result type" box, select "f2" and press the "Show selected result" button. Figure 11 will be shown in the plot tab. It can be seen that the inbreeding coefficient of subpopulation 2 will initially decrease, since its inbreeding coefficient is relatively high from the start, and its individuals are unrelated to the immigrants that enter into this subpopulation the first few generations. This, happens because the initial kinship coefficient has been set to 0 between all subpopulations (see Table 2).

Now, choose "Effective size" from the "Choose result type" box, select "Subpop 1, instantaneous", "Subpop 2, instantaneous", "Subpop 3, instantaneous" and "NeE". Press the "Show selected result button" to show Figure 12 in the "Plot"
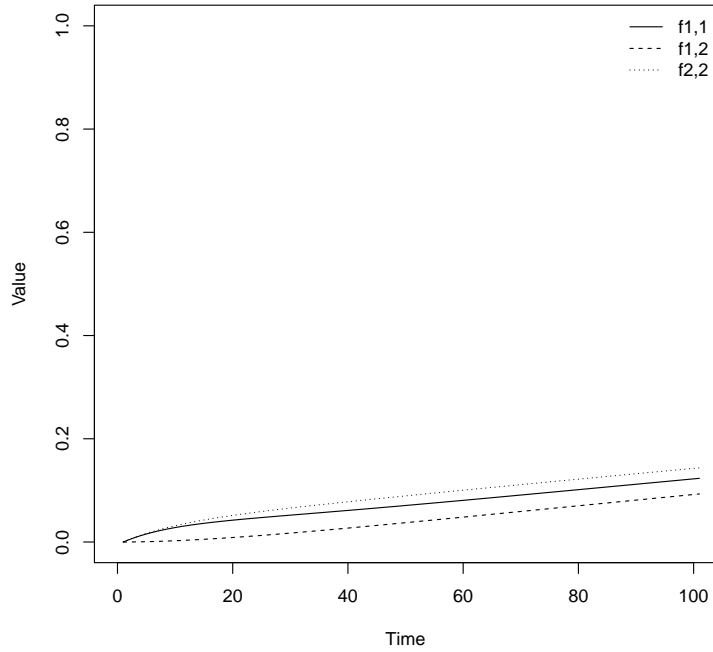
20

Figure 9: Kinship coefficients for subpopulations 1 and 2 of the source sink example.

Table 3: Migration matrix of the complicated migration patterns example.

	To subpop 1	To subpop 2	To subpop 3	To subpop 4	To subpop 5
From subpop 1	45	0	0	0	0
From subpop 2	2	90	1	5	3
From subpop 3	0	4	195	0	1
From subpop 4	3	4	2	390	10
From subpop 5	0	2	2	5	186
Sum	50	100	200	400	200

tab. Since the inbreeding coefficients of subpopulations 1 and 2 decrease the first few generations, their local inbreeding effective population sizes are initially undefined. It can also be seen that the increase of the inbreeding coefficient of subpopulation 3 is small the first generations, so that its local inbreeding effective population size becomes large. In order to see what happens at later generations, mark the "Specify Y-max" check box and set "Y-max" to 2000. Press the "Show selected result button" to show Figure 13 in the "Plot" tab.
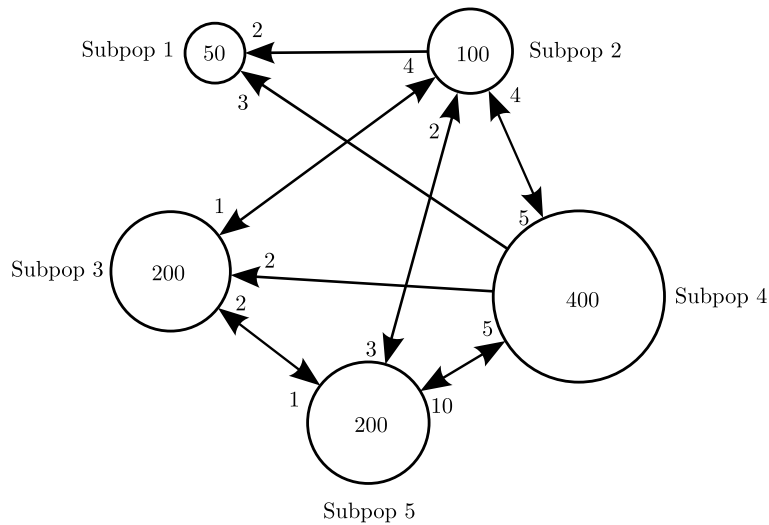
21

Figure 10: Migration scheme (with migration specified as actual numbers) of the complicated migration patterns example.

8.4 Local bottleneck

In this example we start with the model of the previous example. We will model a local bottleneck in subpopulation 2 between generations 10 and 20, so that its size decreases during this time interval. Uncheck the "Constant migration scheme" check-box, and in the "Changes affect" box select "Subsequent generations". Then, change the value "In generation" to 9, set both "Migrants from" and "Migrants to" to "Subpop 2", and let the number of migrants be 20. This will change the migration rate from generation 9 and onward. Now change the value "In generation" to 19 and then change the number of migrants (from subpopulation 2 to itself) back to 90. Switching to the "Population sizes" tab, we see that subpopulation 2 has size 100 between generations 0 and 9, at generation 10 its population size changes to 30 and at generation 20 it changes back to 100 again. The local effective population size will change accordingly. Select "Run analysis" and switch to the result options tab.

Choose "Differentiation" from the "Choose result type" box, select "f1","f2" and press the "Show selected result" button. Then Figure 14 will be shown in the plot tab. By comparing this with Figure 11 we see that the main effect of the bottleneck on the inbreeding coefficient, is a reduction between generations 10 and 20. However, the inbreeding coefficient at generation 100 will be slightly higher with the bottleneck ($f_{t=100,2} = 0.0818$) compared to when the bottleneck is excluded ($f_{t=100,2} = 0.0807$).

8.5 Extinction and recolonization

Start with the model of the previous example. We will model local extinction of subpopulation 2 between generations 10 and 20. When the subpopulation is recolonized, the size will grow until it reaches its previous size of 100. Below we
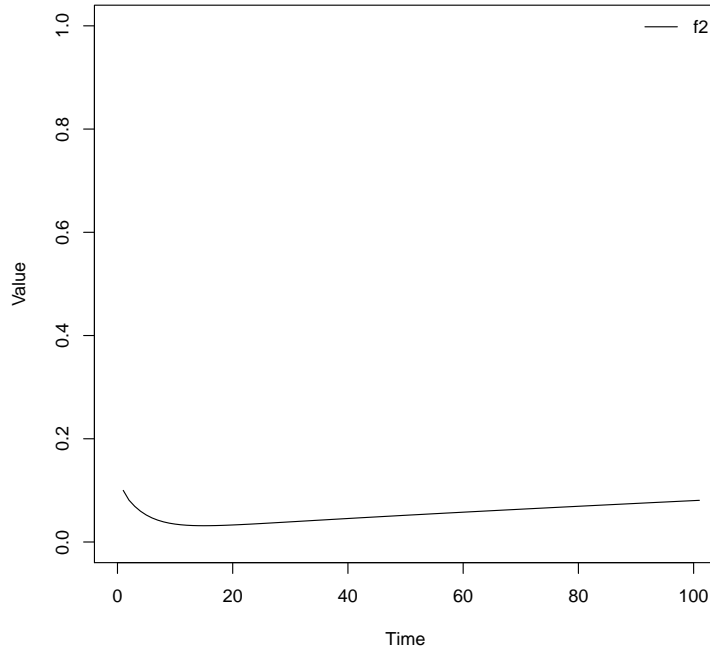
Figure 11: Inbreeding coefficients of subpopulation 2 for the complicated migration pattern example.

will utilize the two possibilities to specify migration. The option "proportions" is useful for population sizes with multiplicative growth or decline. The "actual numbers" option is better to use for e.g. populations in which the number of migrants is should be constant.

Make sure that the "Constant migration scheme" check-box is unchecked, and that the "Changes affect" box is set to "Subsequent generations". Now, change the value of "In generation" to 9, set both "Migrants from" and "Migrants to" to "Subpop 2" and decrease the number of migrants from 20 to 0. Then change migration from subpopulations 3,4 and 5 to subpopulation 2 to 0. Looking at the "Population sizes" tab, the local size of subpopulation 2 should be 0 from generation 10. We also notice that the sizes of the other subpopulations decrease, since they no longer receive immigrants from subpopulation 2. Change the value "In generation" to 19 and modify the number of migrants from subpopulation 3,4 and 5 to subpopulation 2 according to the values of Table 3. Then adjust the value "In generation" to 20 and change the number of migrants from subpopulation 2 to the other subpopulations according to the values of Table 3. In order to increase the size of subpopulation 2 at each generation, change the "Migration specified as" option to "Proportions" and set the migration from subpopulation 2 to itself to 1. This means that, at each generation after generation 20, the individuals in subpopulation 2 will stay in the subpop-
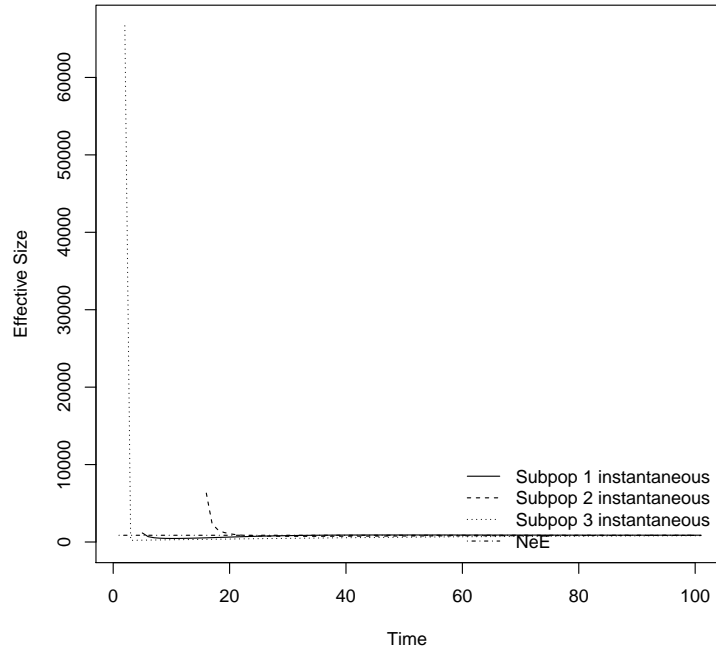
23

Figure 12: Inbreeding effective sizes of subpopulations 1-3, and the eigenvalue effective size of the metapopulation, for the complicated migration pattern example.

ulation. Since it still receives immigrants from the other subpopulations it will grow by 10 individuals per generation. Finally, "In generation" to 29, change the "Migration specified as" option to "Actual numbers" and the number of migrants from subpopulation 2 to itself to 90. If we look at the "Population sizes" tab we should see that subpopulation 2 has reached its original size of 100 at generation 29, and that its size after this generation is constant.

After running this new model, choose "Differentiation" from the "Choose result type" box, select "f1","f2" and press the "Show selected result" button. Then Figure 15 will be shown in the plot tab.
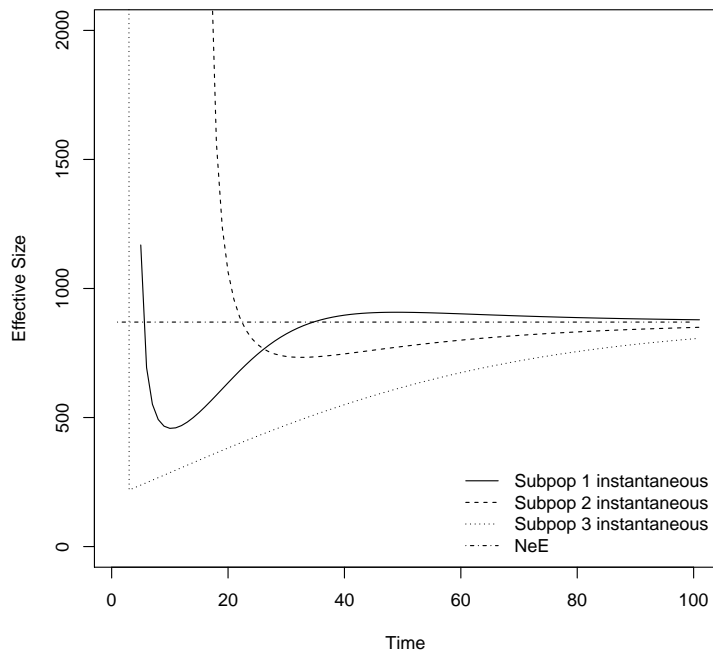
Figure 13: Inbreeding effective sizes of subpopulations 1-3, and the eigenvalue effective size of the metapopulation, with adjusted y-axis, for the complicated migration pattern example.
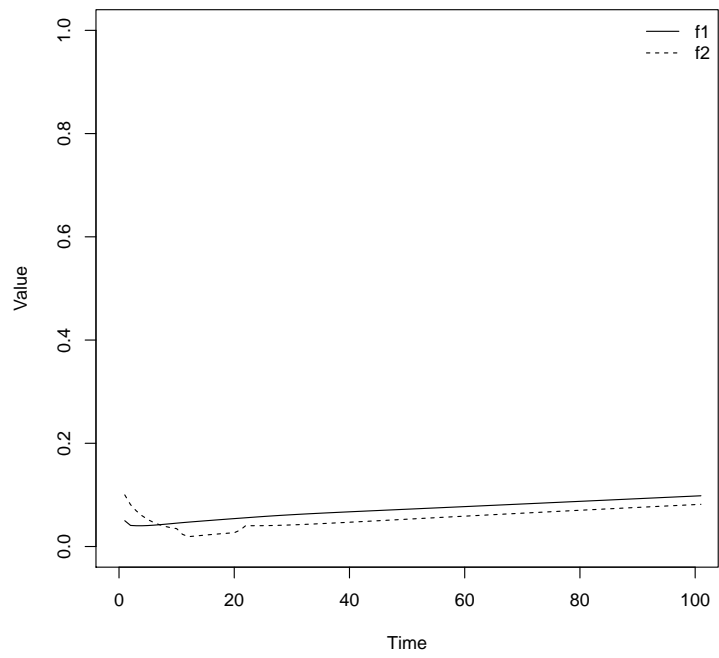
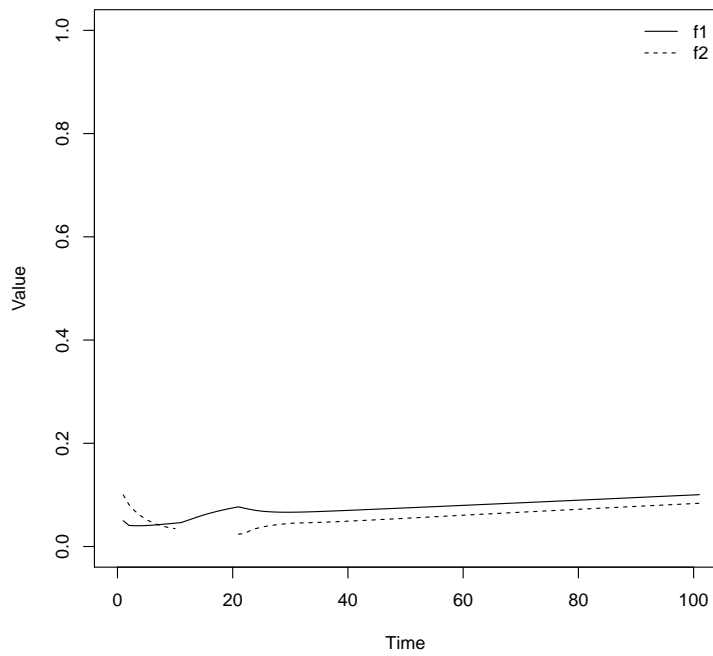Figure 14: Inbreeding coefficients of subpopulations 1 and 2 for the bottleneck example.

Figure 15: Inbreeding coefficients of subpopulations 1 and 2 for the extinction and recolonization example.

9   References

Hössjer O, Olsson F, Laikre L, Ryman N (2014) A new general analytical approach for modeling patterns of genetic differentiation and effective size of subdivided populations over time. *Mathematical Biosciences* **258**, 113-133. DOI: 10.1016/j.mbs.2014.10.001.

Hössjer O, Olsson F, Laikre L, Ryman N (2015) Metapopulation Inbreeding Dynamics, Effective Size and Subpopulation Differentiation - a General Analytical Approach for Diploid Organisms. *Theoretical Population Biology* **102**, 40-59. DOI:10.1016/j.tpb.2015.03.006.

Laikre L, Olsson F, Jansson E, Hössjer O, Ryman N (2016) Metapopulation effective size and conservation genetic goals for the Fennoscandian wolf (Canis lupus) population. *Heredity*. DOI:10.1038/hdy.2016.44

Hedrick, PW (2005) A standardized genetic differentiation measure. *Evolution* **59**(8), 1633-1638.

Meirmans, PG, Hedrick, PW (2011) Assessing population structure: FST and related measures. *Molecular Ecology Resources* **11**, 5-18.

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

Nei, M (1977) *F*-statistics and analysis of gene diversity in subdivided populations. *Ann. Hum. Genet.* **41**, 225-233.