Title:

NeOGen v1.3.0.6.a1 User Manual v1.1.2

Authors:

Dean C Blower[1], Cynthia Riginos[1], & Jennifer Ovenden[1,2]

Affiliations:

School of Biological Sciences[1], Molecular Fisheries Laboratory[2], and School of Biomedical Sciences[3], The University of Queensland, St. Lucia, QLD 4072, Australia.

Corresponding Author:

Dean Blower. The University of Queensland, St. Lucia, QLD 4072, Australia.
dean.blower@uqconnect.edu.au.

DRAFT & CONFIDENTIAL

DRAFT & CONFIDENTIAL – NOT FOR DISTRIBUTION

NEOGEN & NEOGEN-POWER© 2018 DEAN BLOWER ALL RIGHTS RESERVED

NeOGen v1.3.0.6.a1 User Manual v1.1.2

Welcome to……

NeOGen

By

Dean C Blower[1] (dean.blower@uqconnect.edu.au),

Cynthia Riginos[1], & Jennifer Ovenden[1,2]

School of Biological Sciences[1], Molecular Fisheries Laboratory[2], and School of Biomedical Sciences[3], The University of Queensland, St. Lucia, QLD 4072, Australia.

DRAFT & CONFIDENTIAL

Back < Contents > Next NeOGen v1.3.0.6.a1

Contents

DRAFT & CONFIDENTIAL

DRAFT & CONFIDENTIAL

Back < Contents > Next NeOGen v1.3.0.6.a1

Introduction

NeOGen (genetic N_e for Overlapping Generations analysis; pronounced Neo-gen) is point-&-click genetic effective population size (N_e) and demographic modelling software. The software, initially designed for sharks, is applicable to many overlapping-generation species with moderate to low population sizes and fecundity (such as mammals). NeOGen provides a cohesive linkage disequilibrium genetic N_e estimate ($N_{e.LD}$) analysis and population simulation framework that simulates both the demography and genetic composition of a population, given a small subset of fundamental species-specific life-history, demographic and genetic priors. NeOGen then guides the researchers in-silico to establish a tractable sampling regime, resulting in the numbers of samples

DRAFT & CONFIDENTIAL

and loci required to give accurate and precise $N_{e.LD}$ estimates when empirical sampling is performed in the field.

This manual describes the use of the NeOGen graphical user interface (GUI). For details regarding the analyses used by NeOGen and the interpretation of results, please see Blower, Riginos, & Ovenden (submitted): NeOGen: a tool to predict genetic effective population size (N_e) for low-fecundity species with generational overlap, and to assist empirical N_e study design.

Back < Contents > Next NeOGen v1.3.0.6.a1

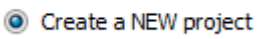
Page 1 - Create, clone, or delete a Project or a Scenario

Page 1 allows for the management of new or existing projects and scenarios. Both projects and scenarios can be created, cloned, or deleted from this page.

Project management

A Project is the container which neatly encapsulates all the Scenario and Sampling Strategies for a given investigation. We suggest that the Project is synonymous with a species and so a species name may be associated with a project. However, researchers may use the project container to logically separate any investigation.

How to - Create a new project

A new Project can be created by selecting the create new project radio button 

and entering a Project Name and a species name. The user is then directed to either create a new Scenario or clone an existing scenario.

Scenario management

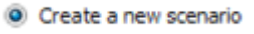
A Scenario is a container used to hold a definition of a species and its population as characterised by the life-history, demographic, and genetic parameters specified by a researcher for a virtual "species". A Scenario always belongs to a project and must be created and named before any species parameters may be entered. Given the uncertainty often associated with species life-history parameters e.g. age at first maturity may be at 6, 7, or 8 years of age, multiple Scenarios may be created and simulated by the researcher. NeOGen allows for multiple Scenarios within a given Project and facilitates the testing of similar scenarios by allowing the 'cloning' of any scenario from any project. Cloning a scenario produces a new scenario with identical species parameters and sampling strategies as the template scenario. This allows the researcher to easily generate multiple

DRAFT & CONFIDENTIAL

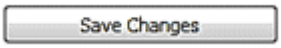
scenarios with small parameter differences and observe to effects of these small differences on proposed sampling strategies.

How to - create a new Scenario

After selecting either an existing Project or a new Project (and supplying a project and species name), a new Scenario may be created by selecting the "Create a new scenario" radio button



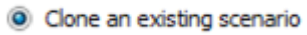
, then entering a name for the Scenario, and clicking Save Changes



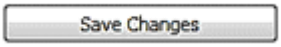
.

How to - Clone an existing Scenario

After selecting either an existing Project or a new Project (and supplying a project and species name), an existing Scenario belonging to this or another project may be cloned all the parameters associated with the chosen Scenario will be saved to a new Scenario under the current Project. This is achieved by selecting the "Clone and existing Scenario" radio button



. This will reveal combo-boxes from which the user can then select the desired project and a scenario to clone from, enter a name for the Scenario, press Save Changes



to complete the cloning process.

Back < Contents > Next NeOGen v1.3.0.6.a1

Page 2 - Specify or modify the current species & population Scenario

Page 2 is where a researcher will specify the Scenario parameters associated with a specific species and its population. NeOGen is not yet capable of sex-specific parameter for just one sex should be used, and using female parameters is recommended.

There are three major categories of priors that must be entered for a Scenario: 1) Life-history parameters, 2) Demographic parameters, and 3) Genetic parameters. All the variables for each of these categories must be supplied before a Scenario can be simulated.

Scenario specification and modification

Life-history parameters (Page 2, Tab 1)

The life-history variables define the age and reproduction properties of a species which are sub-divided into aging properties and litter size properties.

DRAFT & CONFIDENTIAL

Life-history - Aging parameters

Here a researcher may specify the maximum age and ages in between which the species is capable of reproduction.

Maximum age

The maximum age specifies the longevity of an individual. At max age individuals all die.

Maximum mating age

This variable specifies the age that a mature individual is no longer reproductively fertile or competent. The age difference between the maximum mating age and the maximum age represent a senescent life-stage during which individuals are still part of the population but unable to reproduce. This life-stage is best known for humans, but is also known for other species. When a senescent life-stage is not indicated for the species of interest the maximum mating age should be set to the same value as the maximum age.

Minimum mating age

The minimum mating age indicates the age after which the species becomes sexually competent, able to mate and produce offspring. This may not be clearly defined for a species and multiple Scenarios may be required for a researcher to estimate the effect of different minimum mating ages on the effective population size.

Offspring per litter

The number of offspring produced from a mating event i.e. litter size, can be specified with parameters that describe the distribution of offspring resulting from a mating event involving all the adults capable of reproduction. The number of offspring (litter size) yielded by a mated pair may be an absolute number resulting in an approximately fixed number of offspring per parent (ABSOLUTE litter distribution). Alternatively, a probability function (POISSON, BINOMIAL, UNIFORM, GEOMETRIC) may be specified that results in an offspring per litter distribution (that is zero-truncated, i.e. litters of zero offspring are ignored) spanning the entire mating event.

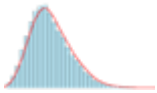
There are 5 offspring per litter distributions to choose from:

(1) *ABSOLUTE Litter size*: The absolute number of offspring per litter i.e. A single number specifying the exact number of offspring in each and every litter

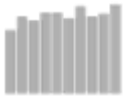
(2) *POISSON Mean litter size*: A Poisson offspring per litter distribution in which case the mean number of offspring per litter will be equal to the variance.

DRAFT & CONFIDENTIAL

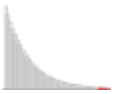
(3) *BINOMIAL Mean litter size*: A Binomial offspring per litter distribution. The mean and standard deviation describe an offspring per litter binomial bell-curve that peaks at the mean value and decays on each side of the mean proportional to the standard deviation.



(4) *UNIFORM Litter size range from a minimum to maximum litter size*: A Uniform distribution of offspring per litter. This distribution results in an equal probability of litters sizes in the range between the minimum and maximum values specified by the researcher.



(5) *GEOMETRIC Mean litter size*: A Geometric offspring per litter distribution. The mean value dictates a geometric curve that peaks approaching 1 offspring per litter and reduces proportionally for each litter size larger than one, which cumulatively gives the specified mean number of offspring per litter.



Demography (Page 2, Tab 2)

The demographic parameters describe the population as a whole and are sub-divided into population size and population mortality rate at each age for each sex.

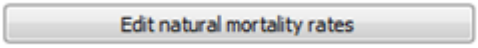
Population size

This parameter is where an approximate estimate of the population size (*N*: total population size, a count of all individuals comprising a finite closed population) may be specified by the researcher. This value may be informed by previous studies, such as mark-&-recapture studies that have estimated population size. However, as this may be largely unknown for many species the researcher may wish to run multiple scenarios which encompass a board range of population size estimates and assess the most likely scenarios by other means.

Distribution of mortality by age & sex

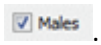
Natural mortality influences the demographic and genetic makeup of a population and thereby the genetic effective population size. For populations of species with significant economic value the rates of natural mortality are often measured or predicted. Other species have relatively unknown rates of mortality may still conform to the survivorship curves predicted by the reproductive behaviour of each species. Editing the natural mortality distribution allows for either a highly precise mortality distribution or a broadly defined distribution.

DRAFT & CONFIDENTIAL

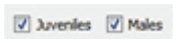
Clicking on the "Edit mortality distribution" button 

opens the mortality distribution screen. The researcher is presented with a range of sliders that represent all age cohorts up to the user-specified maximum age, for both sexes. There are 3 ways that the researcher can enter mortality rates:

1) Manually editing each age cohort mortality proportion individually: The researcher can manually enter the proportion of mortality experienced by each age cohort by adjusting the slider or entering the value as a decimal from 0 to 1. E.g. for a mortality proportion that culls 50% of a particular age cohort enter 0.500 into the appropriate spin-box.

2) Applying a single value to a group: A single mortality rate can be propagated across a group of age cohorts automatically. Firstly, select a group of cohorts by ticking the appropriate sex checkbox e.g. 

. This then enables the user to either manually tick each checkbox of a group of age cohorts individually 

or select an age cohort group of juveniles e.g. immature males 

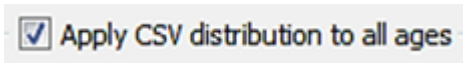
or adults e.g. mature females 

by ticking both the sex and age cohort group checkboxes. A mortality rate value can be applied to all the ticked cohort by selecting "apply value to age groups", entering the desired value in the associated spin-box e.g. 

and clicking the apply 

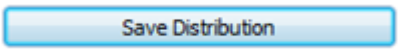
button.

3) Pasting and applying a tab-separated list: The entire mortality rate distribution may be applied en masse by specifying it as a list. Firstly, an age cohort/mortality rate list must be created. There are two ways to do this. Firstly, by specifying two vertical columns in MS Excel (column 1: age in years starting at 0 and ending at maximum age - 1, column 2: mortality rate as decimal value between 0 and 1). The other way is to specify the same columns in a tab-separated list within a text editor.

Once a list has been created, tick the "Apply CSV distribution to all ages" 

and paste the list into the activated text box. Clicking the apply 

button will propagate the list to the entire distribution for the selected sex.

To complete the mortality rate distribution, click the "save distribution" 

button and click the X 
to close the window.

DRAFT & CONFIDENTIAL

Genetic parameters (Page 2, Tab 3)

The genetic component of individuals is specified in this tab. Each parameter in conjunction dictates the genome that each individual will carry and the maximum number of loci that can be interrogated by a researcher. The genome to be simulated has 3 attributes: 1) the number of loci that can be interrogated, 2) the number of alleles per locus or a distribution that represents the number of loci.

Number of loci

The researcher can use the slider or type in the total number of loci that will comprise the simulated genome of each individual. Selecting a low number of loci (e.g. 1,000 for bialleleic loci or 100 for multialleleic loci) is a good starting point as the speed of the population simulation is significantly slowed with increased number of loci.

Alleles per locus distribution

Here the number of alleles (i.e. gene variants per genetic locus) is specified. This can be a Uniform distribution with all loci in a genome having exactly the same number of alleles, or a Binomial distribution of alleles across all loci in the individual's genome. In the latter case the mean and standard deviation of a binomial distribution dictate the number of alleles found at any single locus in the genome with the majority having the mean or similar numbers of alleles and the rarity of loci with alleles much less or much more than the mean being directed by the specified standard deviation. Researchers with experience in developing microsatellite markers for non-model species will recognise this distribution of alleles per loci as one that commonly emerges when randomly screening large numbers of loci for their efficacy as genetic markers. For biallelic loci, the researcher should choose just two alleles per locus (and standard deviation = 0.0, if a binomial distribution is selected).

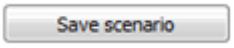
Allele frequency distribution

The allele frequency distribution describes the probability of encountering any particular allele at a particular genetic locus. Typically, a genetic locus will have several common alleles and a lesser number of rarer alleles. This situation is simulated using a Dirichlet randomisation of allele frequencies that is acknowledged as a relatively realistic representation of the variety of allele frequencies at a typical genetic locus. Alternatively, the researcher can choose a non-random distribution of allele frequencies for which alleles have equal probability of occurring for a particular locus, when alleles per locus are UNIFORM, or have a binomial probability distribution if alleles per locus are BINOMIAL. The third alternative is for a researcher to supply empirical pilot locus data from a data file. The AllAlleleFrequencies option allows the user to upload an Arlequin-generated AllAlleleFrequencies text file (Excoffier & Lischer 2010) which automatically provides all the required locus details.

DRAFT & CONFIDENTIAL

It should be noted that allele frequencies will alter with the multiple rounds of simulated mating. Rare alleles are lost by chance (termed genetic drift) after random mating, an effect that increases as population sizes get smaller.

Save Scenario

A scenario can be saved at any time by clicking the Save Scenario button 

, and it is recommended that the user save changes frequently to avoid inadvertent data loss.

Back < Contents > Next NeOGen v1.3.0.6.a1

Page 3 - Run or re-run the current Scenario

Page 3 is where a researcher will specify the final parameters specifying the length of a Scenario run, and can start and monitor a Scenario run. The length of a Scenario depends on 3 factors: 1) the number of independent replicates required, 2) the length of the burn-in period required to equilibrate the demography and genetics prior to 3), the length of the temporal evolution period when the population may be sampled for analysis.

Run or re-run the Scenario

How long will the simulation run (Page 3)?

Number of population replicates

This parameter dictates the number of independent times a Scenario population will be repeatedly simulated. Increasing the replicates adds statistical power to the subsequent sampling strategy LDNe analyses. Increasing population replicates significantly increase the simulation running time. Researchers should start with low numbers of replicates (3 - 10) whilst getting a feel for the data and later perform, longer but more conclusive runs, with increased replicates (20 - 30 or greater; greater than 30 population replicates may take a long while, depending on the population size and the number of loci specified).

Burn-in length (annual matings)

A burn-in simulation period is required to allow the demographic and genetic models to reach equilibrium prior to sampling during the next phase. The minimum length of this period should be 2 lifespans (maximum age x 2) but increasing this by lifespan multiples will ensure equilibrium is reached prior to sampling for analytical purposes.

DRAFT & CONFIDENTIAL

Temporal evolution length (annual matings)

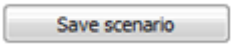
After burn-in, the temporal evolution phase continues with annual mating's but data gathering commences and populations are saved for sampling strategy analysis. Providing that sufficient burn-in has stabilised the population the temporal evolution length need be no greater than 2 lifespans.

Run the Scenario (Page 3)

Select the output path

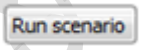
The simulation process creates multiple files which are subsequently referred to by the sampling strategy analysis process. The place where these are stored is important and the default location, which is automatically supplied when a new Scenario is created, is the recommended choice. However, when the researcher needs to locate the Scenario simulation files to manage disk space for example, this option can be changed by clicking the tool-button  and specifying a new simulation output path.

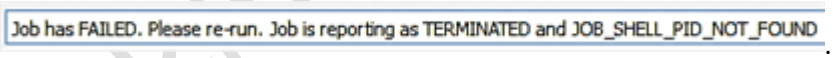
Save Scenario

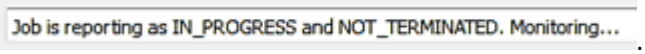
A scenario can be saved at any time by clicking the Save Scenario button 

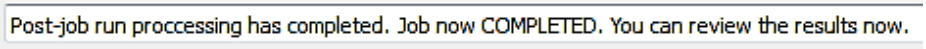
located on both page 2 and page 3. Any Scenario changes must be saved prior to running a Scenario and the user is prompted to do so when run Scenario is pressed

Run Scenario

Once Scenario parameters have been supplied and saved the researcher can run the Scenario by clicking the run Scenario button 

. The user is prompted to confirm, and the Scenario simulation run is initiated in a separate "console" window. The user should avoid closing this console window until the run is finished, as doing so will result in a failed run e.g. 

. The text box next to the run Scenario window provides the user with updates on the progress of the simulation. Whilst a simulation is running the text box regularly refresh a "Job in progress. Monitoring…" message 

. The user can be assured that the run is finished when the "Run completed" message 

is displayed. When the run has completed the user can close the console window by clicking on the X or selecting the window and pressing any key.

DRAFT & CONFIDENTIAL

Back < Contents > Next NeOGen v1.3.0.6.a1

Page 4 - Review the Scenario results

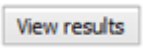
Page 4 presents the results of the Scenario simulation. Here the researcher can review the demographic profile produced by the Scenario simulation and note two things. Firstly, is this a reasonable demographic representation of the target species? Secondly, which age cohorts are to be the target of sampling and will their size support the sampling intensity that the researcher proposes?

Review the Scenario results

Scenario demographic results (Page 4)

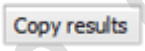
The Scenario demographic results are displayed as a plot in preview size. This is provided to give the researcher a quick overview of the results. A closer look at this plot is achieved by clicking View results or by inspecting the plots after copying them to a location of the researcher's choice.

View results

The push button "View results" 

generates a PDF document of the Scenario demographic results that may be saved to a location of the users choice.

Copy results

The button "Copy results" 

allows the user the results plot image file to a destination of their choice.

For details regarding the analyses used by NeOGen and the interpretation of results, please see Blower, Riginos, & Ovenden (submitted): NeOGen: a tool to predict genetic effective population size (N_e) for low-fecundity species with generational overlap, and to assist empirical N_e study design.

Back < Contents > Next NeOGen v1.3.0.6.a1

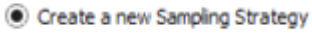
Page 5 - Create, clone, or delete a Sampling Strategy

Page 5 allows for the management of new or existing Sampling Strategies. Sampling strategies can be created, cloned, or deleted from this page. A Sampling Strategy container holds all the parameters required to assess the combined power of samples and loci to make a reasonable linkage disequilibrium genetic N_e estimate ($N_{e.LD}$) using the LDNe method (Waples & Do 2008) implemented by NeEstimator v2.01 (Do *et al.* 2014).

DRAFT & CONFIDENTIAL

Sampling Strategy management

How to - create a new Sampling Strategy

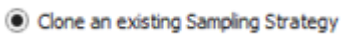
A new Sampling Strategy may be created by selecting the "Create a new Sampling Strategy" radio button 

, entering a name for the Sampling Strategy (or accepting the default name), and clicking Save Changes 

.

How to - Clone an existing Sampling Strategy

An existing Scenario belonging to this project may be cloned and all the parameters associated with the chosen Sampling Strategy will be saved to the new Sampling Strategy under the current Scenario.

This is achieved by selecting the "Clone an existing Sampling Strategy" radio button 

. This will reveal a Scenario combo-box from which the user can then select the desired Scenario to clone from. Enter a name for the Sampling strategy (or accept the default name), and press Save Changes 

to complete the cloning process. Cloning a Sampling Strategy from a different project cannot be accomplished here, but can be achieved by cloning the Scenario from the Project which contains the desired Sampling Strategy and modifying both the new Scenario and Sampling Strategy to suit.

Back < Contents > Next NeOGen v1.3.0.6.a1

Page 6 - Specify or modify the current Sampling Strategy

Page 6 allows for the modification of the current Sampling Strategy. Two main parameters are required: 1) the number of samples a researcher may feasibly gather, and 2) the number of loci a researcher may feasibly obtain.

Sampling Strategy specification

Sample size range (Page 6, Group 1)

Here the researcher chooses the sample size range that will form the basis of the current Sampling Strategy analysis

Maximum samples

Here the research specifies the maximum number of samples that they feel can feasibly be obtained for the proposed research project

DRAFT & CONFIDENTIAL

Increment

This parameter controls the range and number of sample size estimates that will be made, e.g. if a researcher wishes to see the analysis performed with 400 samples down to 100 samples and increment of 100 will produce $N_{e.LD}$ estimates for 100, 200, 300, & 400 samples. However, as the Sampling Strategy analysis combines samples sizes and locus quantities up to a maximum of 12 possible combinations, the minimum samples may be adjusted according the existing locus quantity parameters. Using the previous example and assuming that 3 locus combinations are already specified, halving the sample increment to 50 will not produce 6 sample estimates from 100 - 400 but to preserve the total sample/locus combination at or below 12 (4 sample size x 3 locus quantities) the sample minimum will be automatically adjusted upwards to 250 again giving 4 sample sizes of 250, 300, 350, & 400. Here, only reducing the locus quantity combination will allow the researcher to increase the sample sizes included in the analysis. This process can be examined visually by observing the results displayed in the "Check your sample size and locus range combination" group box when changing sample size and locus range parameters.

Minimum samples

As mentioned above, this parameter is set automatically dependent on the combination of sample sizes and loci numbers chosen by the researcher.

Locus quantity range (Page 6, Group 2)

Here the researcher chooses the locus quantity range that will form the basis of the current Sampling Strategy analysis

Maximum loci

Here the research specifies the maximum number of informative genetic loci that they feel can feasibly be generated or are available for the proposed research project

Increment

This parameter controls the range and number of locus quantity estimates that will be made, e.g. if a researcher wishes to see the analysis performed with 30 loci down to 10 loci an increment of 10 will produce $N_{e.LD}$ estimates for 10, 20, & 30 loci. However, as the Sampling Strategy analysis combines samples sizes and locus quantities up to a maximum of 12 possible combinations, the minimum loci may be adjusted according the existing sample size parameters (as explained for sample size increment previously).

Minimum loci

As mentioned above, this parameter is set automatically dependent on the combination of locus quantities and sample sizes chosen by the researcher.

DRAFT & CONFIDENTIAL

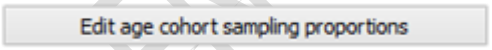
Check your sample size and locus range combination (Page 6, Group 3)

This group box is for display only and allows the user to inspect the sample size and locus quantity combinations that will be analysed and displayed in the Sampling strategy results. Changing the sample size and locus quantity parameters will automatically update this display and the researcher may experiment with the combination until satisfied that the desired sample size/ locus quantity combination analyses (up to a maximum of 12 combinations, e.g. 4 sample size x 3 locus quantities = 12 combinations) will be performed.

Specify the proportion of samples for each age cohort (Page 6, Group 4)

This is where the researcher will specify the sampling proportion for each age cohort with reference to the Scenario demographic simulation results.

Edit age cohort sampling proportions

Click the "Edit age cohort sampling proportions" button 

to open the window for editing.

Sampling proportions by age (Page 6, Sampling proportions window)

In the "Sampling proportions by age" window the researcher can specify, to the best of their knowledge, the most feasible empirical sampling regime. From pilot studies a researcher may have a feel for the age cohorts that can be realistically sampled and the likelihood of obtaining a particular age cohort or a group of cohorts at a particular stage of maturation. For example, white shark adults are rare and particularly difficult to sample. Similarly, neonate white sharks are hard to locate. However, the location of sub-adult aggregations are known and are relatively easily sampled, making the juvenile cohorts (excluding neonates) to recently matured the most feasible target of sampling. In this case, the researcher would tick the age cohorts around the age at first maturity and give a sampling proportion as indicated by pilot sampling and the simulated Scenario demographic profile.

This window presents a range of sliders that represent all age cohorts up to the user-specified maximum age. However, the first decision is to select a sampling distribution type.

Sampling distribution type (Page 6, Sampling proportions window, Group 1)

The depth of knowledge a researcher may have regarding the demography and intensity of sampling that can be considered for their target species. The sampling distribution type allows for either a detailed sampling regime and a more broadly defined sampling regime.

USER_PROPORTIONS Sampling distribution type: This should be chosen if the researcher will supply the proportions for each chosen age cohort.

DRAFT & CONFIDENTIAL

USER_AGE_COHORTS Sampling distribution type: This should be chosen if the researcher is comfortable just select the desired age cohorts and then allows the program to calculate the sampling proportions based on the maximum sample size and the size of each age cohort as indicated by the Scenario simulated demography.

When the *USER_PROPORTIONS* Sampling distribution is selected this window operates in a very similar way to the "Mortality by age and sex" window for a Scenario and, excluding the sex specification, the data can be entered in the same ways. There are 3 ways that the researcher can enter sampling proportions:

1) Manually editing each age cohort sampling proportion individually.

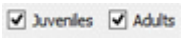
2) Applying a single value to a group: A single sampling rate can be propagated across a group of age cohorts automatically

3) Pasting and applying a tab-separated list: The entire sampling rate distribution may be applied *en masse* by specifying it as a list. However, an age cohort/sampling proportion list must be created beforehand.

If ignoring the need to choose a sex, each of these actions 1 - 3 can be performed in the same way as editing mortality rates.

When the *USER_PROPORTIONS* Sampling distribution is selected the actions of the user are restricted to ticking the age cohorts that should be sampled in 3 ways:

1) Ticking the desired age cohorts one-by-one: When USER_PROPORTIONS is selected the age cohort tick boxes are enabled and may be selected ad hoc.

2) Selecting a grouping by maturity: For example, ticking juveniles and/or adults 

(and then editing manually one-by-one if required)

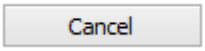
3) Pasting and applying a tab-separated list: The entire sampling rate distribution may be applied *en masse* by specifying it as a list. However, an age cohort/sampling proportion list must be created beforehand with the proportion of the desired age cohorts listed as 1.00.

Actions 1 - 2 are self-explanatory, but action 3, can be performed in the same way as editing mortality rates, ignoring the specification of a sex.

DRAFT & CONFIDENTIAL

To complete the sampling proportion distribution, click the "save distribution"

button and click "Cancel" button



or the X

to close the window.

Back < Contents > Next NeOGen v1.3.0.6.a1

Page 7 - Run or re-run the current Sampling Strategy

Page 7 is where a researcher will specify the final parameters of a Sampling Strategy run, and can start and monitor a Sampling Strategy run. The length of a Sampling Strategy run depends on 3 factors: 1) the size of the sample combinations, 2) the number of loci requested for analysis, and 3) the number of independent LDNe replicates selected by the researcher.

Run or re-run the Sampling Strategy

Sampling Strategy run parameters (Page 3)

LDNE Pcrit

This parameter is required by the linkage-disequilibrium method of $N_{e.LD}$ estimation and specifies the allele frequency below which alleles are excluded from the resulting LDNe estimate. Typically, when sample size is greater than 20 then Pcrit 0.02 or 0.05 is the best choice for balancing the loss of power by excluding alleles with the downward bias exerted by low frequency (rare) alleles (Waples & Do 2010).

Number of LDNe replicates

This parameter dictates the number of independent times a random sample is taken and evaluated for $N_{e.LD}$ using the LDNe method. Increasing the replicates adds statistical power to the subsequent sampling strategy $N_{e.LD}$ analyses but can also increase the simulation running time. Researchers should start with low numbers of replicates (10 - 20) whilst getting a feel for the data and later perform, longer but more conclusive runs, with increased replicates (30 – 50 or greater).

Run the Sampling Strategy (Page 7)

Select the output path

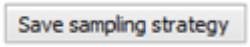
The simulation and sampling strategy process create multiple files, and the default location, which is automatically supplied when new Sampling Strategy is created, is the recommended choice. However, when the researcher needs to relocate the Sampling Strategy files to manage disk space, this option can be changed by clicking the tool-button

and specifying a new output path.

DRAFT & CONFIDENTIAL

Save Sampling Strategy

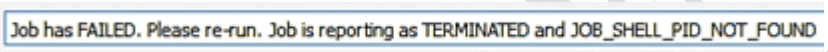
A scenario can be saved at any time by clicking the Save Sampling Strategy button

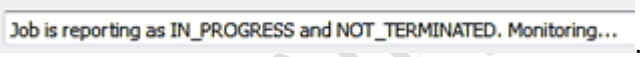


located on both page 6 and page 7. Any Sampling Strategy changes must be saved prior to commencing a run and the user is prompted to do so when run Sampling Strategy is pressed

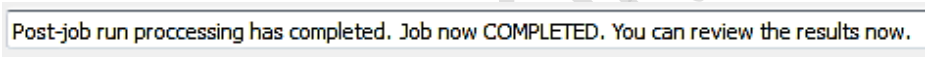
Run Sampling Strategy

Once Sampling Strategy parameters have been supplied and saved the researcher can run the Sampling Strategy by clicking the Run Sampling Strategy button 

. The user is prompted to confirm, and the Sampling Strategy analysis run is initiated in a separate "console" window. The user should avoid closing this console window until the run is finished, as doing so will result in a failed run e.g. 

. The text box next to the run Sampling Strategy window provides the user with updates on the progress of the simulation. Whilst a simulation is running the text box regularly refresh a "Job in progress. Monitoring…" message 

. The user can be assured that the run is finished when the "Run completed" message



is displayed. When the run has completed the user can close the console window by clicking on the X or selecting the window and pressing any key.

Back < Contents > Next NeOGen v1.3.0.6.a1

Page 8 - Review the Sampling Strategy results

Page 8 presents the results of the Sampling Strategy analysis. Here the researcher can review each sample-size specific Sampling Strategy demographic plot and the Sampling Strategy LDNe analysis plot showing the accuracy of each sample size / locus quantity combination.

Review the Sampling Strategy results

Sampling Strategy results (Page 8)

The Sampling Strategy LDNe analysis results are displayed as a plot in preview size. This is provided to give the researcher a quick overview of the results. A closer look at this plot is achieved by clicking View results or by inspecting the plots after copying them to a location of the researcher's choice.

DRAFT & CONFIDENTIAL

View results

All the plots generated for a Sampling Strategy may be previewed by pressing the Previous \ Next buttons  . The push button "View results"  generates a PDF document of the Sampling Strategy results that may be saved to a location of the users choice.

Copy results

The button "Copy result"  allows the user to copy all the results plot image files to a destination of their choice.

For details regarding the analyses used by NeOGen and the interpretation of results, please see Blower, Riginos, & Ovenden (submitted): NeOGen: a tool to predict genetic effective population size (N_e) for low-fecundity species with generational overlap, and to assist empirical Ne study design.

DRAFT & CONFIDENTIAL

References

Blower, D, Riginos, C, & Ovenden, JO (submitted) NeOGen: a tool to predict genetic effective population size (N_e) for low-fecundity species with generational overlap, and to assist empirical N_e study design. *Molecular Ecology Resources.*

Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR (2014) NEESTIMATOR V2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources* **14**, 209-214.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.

Waples RS, Do C (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources* **8**, 753-756.

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications* **3**, 244-262.

DRAFT & CONFIDENTIAL