

# POPSIM: a computer program for simulation of tree breeding programs over multiple generations

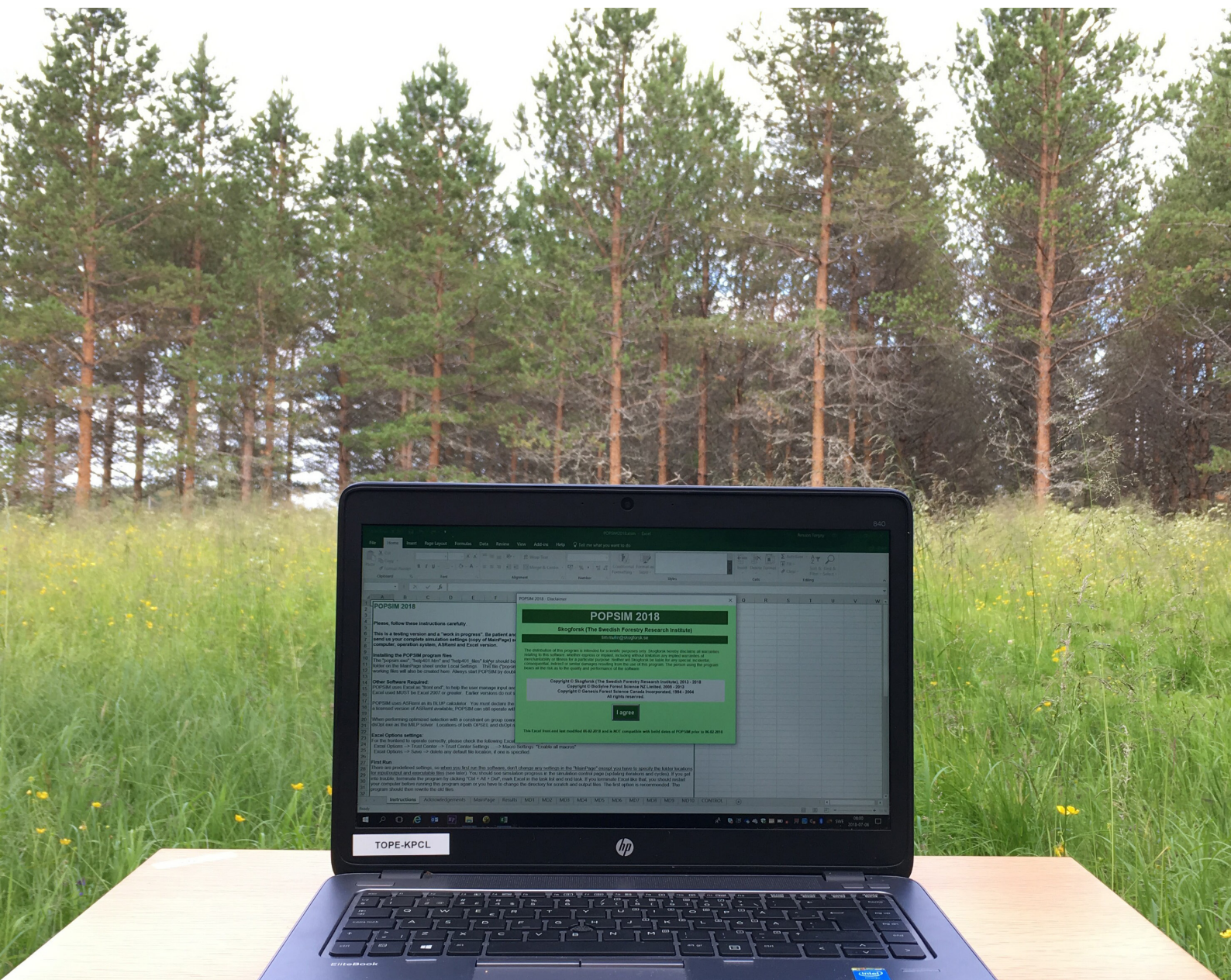


PHOTO: TORGNY PERSSON/SKOGFORSK

## Sammanfattning

Utvärdering och optimering av förädlingsstrategier för skogsträd kan med fördel hanteras med stokastisk simulering. Rekombination av alleler vid reproduktion(korsning) är en stokastisk effekt, och således även uttrycken av genetiska effekter i en population. Denna process kan efterliknas i en simulering, när en sluten population påverkas enligt en förutbestämd förädlingsstrategi. När en simulering upprepas många gånger kan slutsatser dras om sannolikhetsfördelningen av olika parametrar ifrån ett scenario, som hjälper till att bilda en uppfattning om ett ”typiskt” utfall. Få antaganden behövs för att prediktera resultatet av olika förädlingsmetoder med komplicerade släktskap över multipla generationer av förädling.

POPSIM är ett exempel på en parametrisk simulering. Med detta program kan första generationens plusträd (founders) förädlas enligt en given strategi. Programmet kan generera avkommor både för avkommeprövning och för den kommande förädlingspopulationen, utvärdera data ifrån fältförsök (BLUP avelsvärden), selektera fördelaktiga träd med specificerad selektionsmetod, generera produktionspopulationer (tex fröplantager, klonblandningar) och beräkna olika populationsegenskaper som exempelvis den genetiska vinsten, effektiv populationsstorlek och diversitetsförlust. Det finns många möjligheter och alternativ ifrån operativa förädlingsprogram inbyggda i POPSIM, som exempelvis avkommeprövning, flexibla korsningsdesigner som t.e.x. rangordnade och obalanserade, klonupprepning i fältförsök, olika selektionsmetoder inkluderat restriktion på släktskapet hos de utvalda träden.

POPSIM är ett projekt som pågått under 25 år och är under ständig utveckling. Nu har nya, viktiga verktyg lagts till för att anpassa programmet till europeiska behov. Dessa verktyg inkluderar optimerat urval till fröplantage med diversitetsrestriktioner över samtliga generationer, optimerat urval till förädlingspopulationen med lika eller olika stort (obalanserat) bidrag från de selekterade föräldraträden, avancerade förädlings-system med fri avblomning eller kontrollerad pollenblandning (polymix) med begränsningar av diversitetsförlusten, automatisk generering av korsningsschema med rangordnade korsningar med obalanserat föräldrabidrag och restriktion för att närbesläktade träd korsas.

Ett användarvänligt gränssnitt gör det enkelt att ställa upp jämförande scenarion, med hjälp av välkända funktioner i ett excelark. Efter att simuleringen är genomförd kan användaren ladda in resultaten i excelarket för vidare behandling av exempelvis medelvärden och standardavvikelser av alla parametrar för varje generation av förädling. Exempelvis kan förändringar i genetiska effekter och varianser, genetiska vinster i produktionspopulationerna samt inavelsökning och effektiva populationsstorlek följas. Användaren kan efteråt behandla resultaten och göra informativa och lättöverskådliga grafer. POPSIMTM kräver Excel 2007 (eller en senare version) men kan köras på alla 64-bits datorer med Windows. BLUP-beräkningarna kräver att mjukvaruprogrammet ASReml finns tillgängligt (finns hos VSN International). Dock är det inte nödvändigt att använda BLUP som metodik för avelsutvärdering.



skogforsk

Uppsala Science Park, 751 83 Uppsala  
skogforsk@skogforsk.se

skogforsk.se

©Skogforsk 2018 ISSN 1404-305X

## **Preface**

This research has received funding from (1) Föreningen Skogsträdsförädling (The Tree Breeding Association, Sweden), (2) the European Union Horizon 2020 Research and Innovation Program under grant agreement no. 676876 (Project GenTree), and (3) Skogforsk (The Swedish Forestry Research Institute).

Sävar 2018-08-01

Tim J. Mullin



# Contents

Sammanfattning .....	2
Preface .....	3
Summary .....	5
Introduction .....	6
Understanding parametric simulation models .....	7
Understanding the simulation process .....	9
Hardware/software requirements for POPSIM .....	9
Becoming a registered user of POPSIM .....	10
A note on copyright, licensing and distribution .....	10
Distribution of POPSIM .....	10
Installing POPSIM .....	11
Setting up scenarios – the input interface .....	12
Set up scenarios .....	12
Supply simulation control parameters .....	12
Scenario identification: .....	12
Simulation dimensions: .....	13
Genetic structure of the managed population: .....	13
Progeny testing specifications: .....	14
Specifications for generating recruitment populations: .....	16
Selection of the breeding population: .....	18
Optimum selection switches .....	21
Selection of seed orchard production population: .....	22
Orchard selection restrictions .....	23
Selection of a full-sib family production population: .....	23
Selection of a mixture of selected clones: .....	24
Local settings: .....	24
Running a series of scenarios .....	25
Retrieving and viewing output .....	26
Saving scenario parameters and results .....	26
Analysing simulation results .....	26
When things go wrong! .....	27
Help to improve POPSIM .....	27
References .....	28
Appendix I – How to use the mating design grid function in the POPSIM Excel workbook .....	31
Appendix II – How POPSIM calculates “combined index” .....	32
Appendix III – How POPSIM calculates “group merit” .....	33

## Summary

Stochastic simulation is particularly well suited to the evaluation and optimisation of tree breeding strategies. The recombination of alleles during reproduction, and thus sampling of genetic effects, is a “stochastic” process. Simulation can mimic this sampling when a closed population is managed in a specified way. When the simulation is repeated, the distribution of “actual” outcomes reflects the effect of random sampling on the likelihood of achieving the “average” outcome. Few assumptions are necessary, making it possible to predict outcomes for complicated pedigrees and combinations of breeding methods over several cycles of breeding.

POPSIM is an example of a parametric simulation. With it, we can breed founders according to a plan, generate offspring for progeny testing and for recruitment populations, evaluate field test data (BLUP breeding values), make selections according to specified criteria, generate deployment populations, and evaluate their genetic gain, effective size and diversity. Included in POPSIM are many features of “real” breeding programmes, including progeny testing, flexible mating designs, assortative and unbalanced mating, clonal replication of field tests, various approaches to selection, including optimal contributions from selected trees.

While POPSIM has its roots in 25 years of collaborative development, important new features have recently been incorporated to adapt the tool to European requirements. These include optimal selection of seed orchards with a constraint on diversity, optimal selection of breeding populations with either equal or unequal contributions from selected parents, advanced breeding with open pollination, polymix breeding with constraints on diversity, and automatic generation of mating designs that accommodate unequal parent contributions with positive assortative mating, while avoiding a specified level of coancestry.

A user-friendly interface makes it simple to set up scenarios for comparison, using familiar tools in an Excel spreadsheet environment. After the simulations are complete, the user can migrate the results into the Excel workbook and automatically generate a report of means and standard deviations for all parameters over all iterations, and for each cycle of breeding, changes in genetic effects and variances, genetic gain from various production and deployment scenarios, as well as inbreeding and effective population size. The user can then save all settings and tabulated results, and manipulate the data as required with Excel analysis and graphing functions. The POPSIM package requires Excel 2007 (or later), but will run under any 64-bit Windows OS environment. BLUP calculations are available, provided the user has access to a licensed version of ASReml (available from VSN International).

## Introduction

The breeding of forest trees, like most other higher crop plants, is generally carried out as repeated cycles of selection, mating, and testing, within managed populations. Figure 1 gives a schematic representation of such a programme. The process begins with a “recruitment population”, which in the first cycle is usually a wild forest of undomesticated trees. Individuals are selected in the recruitment population to form a “breeding population”, which is often archived by grafting to facilitate management. Individuals in the breeding population are mated and their progeny established in field tests, which become the recruitment population for the next cycle. Meanwhile selected individuals are identified to form “production populations”, such as seed orchards or tissue archives that will produce stock for forest planting. While many changes can occur over time in the genetic structure of the recruitment, breeding and production populations, the impact of a programme must ultimately be measured by the genetic quality and diversity of deployed planting stock.

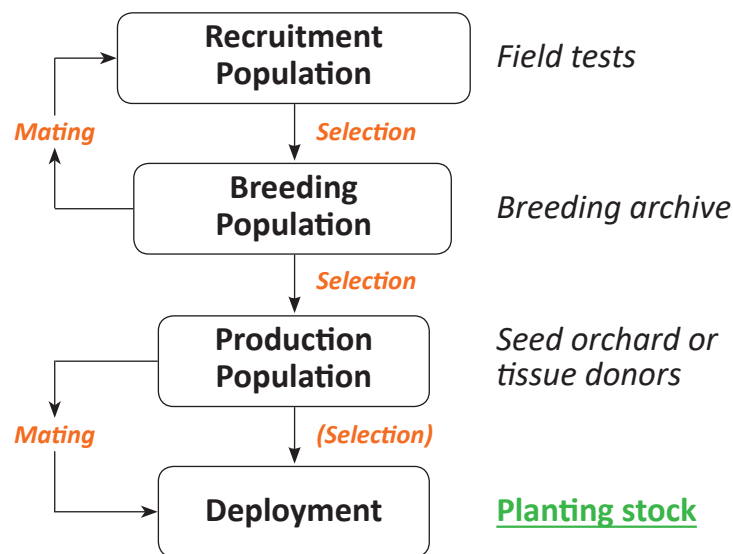


Figure 1. Schematic representation of a typical tree breeding programme.

The prediction of genetic gain from one round of selection in a tree improvement programme is a relatively straightforward calculation, requiring only appropriate estimates of heritability, selection intensity and the phenotypic variance of the selection units. There are many papers in the literature demonstrating how this basic calculation can be applied to a wide variety of situations found in applied tree breeding programmes (Lindgren et al. 2008; Shelbourne 1969, 1992; Verry et al. 2000).

Although such gain calculations are useful, comparisons of breeding strategies based only on gain estimates from one round of selection are often inadequate. More often the objective of an improvement programme will be to maximise gain over time after several cycles of breeding. This objective will often be further constrained by the loss of genetic variance, accumulation of inbreeding, effective population size, and ultimately cost (Mullin, 1994).

Using computers to simulate genetic systems is not really new; the technique has been reported in the literature for at least 35 years (Fraser, 1957). In genetic systems, the random recombination of alleles through sexual reproduction is a process that lends itself to simulation by stochastic methods. By repeating the simulation through several iterations, one can predict the mean outcome of stochastic events and describe the variance of outcomes about this mean.

Our objective was to provide breeders with a generalized decision-support tool to critically examine multi-generation breeding plans. The theoretical basis for the simulation was first documented by Mullin and Park (1995), and has been expanded periodically in other published work where the tool was applied (Andersson et al. 1998; Bouffier et al. 2011; Lindgren and Mullin 1997; 2006; Lstibůrek et al. 2004a, b; 2005; Rosvall et al. 2015; Rosvall et al. 1998; Rosvall and Mullin 2003; Rosvall and Mullin 2013; Rosvall et al. 2003; Weng et al. 2009).

This report describes the application of POPSIM, for the simulation and analysis of tree breeding population management strategies on modern personal computers. The report is an update to earlier documentation (Mullin et al. 2010) and includes details on many upgrades and features added to the software in the last 8 years.

While this documentation is intended to be complete, we have assumed that the user is familiar with the terminology and procedures commonly used in the design and application of tree breeding strategies. Some users, particularly those without formal training in quantitative genetics, may find it useful to review concepts in an introductory textbook on forest tree breeding or crop improvement (e.g., White et al. 2007). Similarly, the guide assumes a general familiarity with basic procedures and configuration of PC-compatible personal computers. Details on these procedures can be found in the original hardware and operating system documentation, and in a host of other references available in any computer store or library.

## Understanding parametric simulation models

Deterministic models predict the outcome of a process through algebraic descriptions of the process and/or regression equations developed from observed sample data. Stochastic simulations attempt to “mimic” critical components of the process that are subject to random variation. The result of every simulation is different, as it is driven by stochastic sampling. To generalise about the outcome of a process, it is necessary to repeat the simulation many times, and observe the mean outcome and the variance around it.

Stochastic simulation models of genetic systems fall into two general categories: (1) allelic models, and (2) parametric models. Allelic models mimic the distribution and sampling of alleles and individual loci affecting a trait, while parametric models describe distribution of genetic effects, based on the infinitesimal quantitative genetic model (Fisher, 1918), where variation in a trait (or index) is assumed to be “controlled by genes at an infinite number of loci, each with an infinitely small effect, in linkage equilibrium”.

POPSIM generally uses a parametric approach. In the model, each genotype is considered to be the sum of independent genetic and environmental effects, and the total phenotypic variation in a population is described as the sum of independent variances for each of

these effects. The additive effect for an individual is generated from the mid-parent family mean and a within-family Mendelian sampling term:

$$A'_i = \frac{A_f + A_m}{2} + r \sqrt{\frac{1 - 0.5(F_f + F_m)}{2}} \sigma_A^2 \quad [1]$$

where  $A'_i$  is the additive effect for progeny  $i$ ;  $A_f$  and  $A_m$  are the additive effects of the female and male parents, respectively;  $F_f$  and  $F_m$  are the corresponding coefficients of inbreeding;  $\sigma_A^2$  is the additive variance; and  $r$  is a normal deviate.

The dominance effect for an individual is generated as:

$$D'_i = r_{fm} \sqrt{\frac{1}{4} \sigma_D^2} + r_i \sqrt{\frac{3(1 - 0.5(F_f + F_m))}{4} \sigma_D^2} \quad [2]$$

where  $\sigma_D^2$  is the dominance variance;  $r_{fm}$  and  $r_i$  are random normal deviates for each family ( $fm$ ) and individual ( $i$ ), respectively.

Inbreeding depression is simulated by reducing the dominance effect by  $bF_{fm}\sigma_p$ , which describes the regression of inbreeding depression on the inbreeding coefficient of the family members  $F_{fm}$ , where the regression coefficient  $b$  expresses the reduction in phenotypic value in units of phenotypic standard deviation,  $\sigma_p$ , for the trait in the unselected base population.

Epistatic variance  $\sigma_I^2$  and effects  $I'_i$  are considered to be within-family, so a simple sampling is made from a normal distribution of epistatic effects:

$$I'_i = r \sqrt{\sigma_I^2} \quad [3]$$

The environmental effect  $E'_{ik}$  for the  $k$ th ramet of the  $i$ th genotype assumes that the environmental variance  $\sigma_E^2$  is normally distributed and equal in all generations:

$$E'_{ik} = r \sqrt{\sigma_E^2} \quad [4]$$

For each of the genetic and environmental effects, estimates of mean and variance are calculated from the generated data for the  $N$  trees in the breeding population. Group coancestry (Cockerham 1967) and inbreeding are calculated from the pedigree. Status number  $N_s$  is an effective number calculated for each generation  $t$  from group coancestry  $\Theta$  (Lindgren et al. 1996):

$$N_{s_t} = \frac{1}{2\Theta_t} \quad [5]$$

By repeating a simulation of a given scenario many times, both the stochastic variation in expected genetic effects and variances of a single scenario can be observed and described by simple statistics.



## Understanding the simulation process

The general approach used by the POPSIM simulator is to generate genetic and environmental effects for a base population of trees that are then mated according to a user-defined breeding plan. Individual progeny from the breeding plan are generated, simulating the recombination of genetic effects and the random environmental effects during field testing. Selection is carried out among the test progeny to assemble various kinds of production populations for deployment of improved stock, and to advance the breeding population to the next cycle of breeding. Genetic variances are calculated directly from the simulated data and reported for each generation, together with the average level of inbreeding and effective population size. The cumulative genetic gain realized by each type of production population is presented as a percentage of the trait mean in the base breeding population.

Of course, a real breeding programme is carried out only once, and while the expected outcome is the central tendency observed from many iterations, the actual outcome of any given iteration can differ. In addition to the mean values for each population parameter, POPSIM also reports the standard deviation, which describes the distribution of single iterations around the mean.

To carry out an optimisation, the user provides parameters that describe the type of population management and the dimensions applicable to the species and programme of interest. Experience or historical data are used to establish realistic genetic parameters and the expected timeline and costs are tabulated. Additional scenarios are described, where parameters are changed systematically, while keeping the overall investment per unit time constant. The results will give insight on the sensitivity of a particular breeding plan to actual genetic and management parameters. By rescaling the scenarios according to different budgets, the simulations can also investigate the sensitivity of breeding plans to available funding.

## Hardware/software requirements for POPSIM

POPSIM is written as a 64-bit program for PC-compatible computers, but will run on any computer under the Microsoft Windows Operating System from version “XP” to present, provided a 64-bit version is installed. While multitasking is possible when running POPSIM, the processor requirements of simulation may result in poor usability, so it is normally recommended to run the program with all other applications closed.

POPSIM requires that the user have a version of Microsoft Excel, no older than version 2007.

If the user wishes to simulate BLUP estimation of breeding values (other options are available), run-time access to a licensed version of ASReml (Gilmour et al. 2009) will be required. POPSIM has been tested with all versions of ASReml up to version 4.

Similarly, if the user wants to specify optimum contributions selection, POPSIM will require access OPSEL version 2.0, together with access to dsOpt.exe, filesocp.exe, and/or ecos-wrapper.exe, depending on the type of selection required (Mullin 2017a). In such cases, the user may also want to generate mating designs for unequal contributions, requiring access to XDesign Version 1.0 (Mullin 2017b).

## Becoming a registered user of POPSIM

POPSIM and its supporting tools are available from Skogforsk at [www.skogforsk.se/popsim](http://www.skogforsk.se/popsim). A download link is provided to supply users with a zipped package of components and examples. While distribution of POPSIM is not restricted, subscription to the users' email list is strongly encouraged. Periodic updates, bug fixes and lists of known issues will be circulated to all subscribers as they are released.

## A note on copyright, licensing and distribution

While copyright for POPSIM is held by Skogforsk, distribution is without fee for research and non-commercial use under GNU General Public License Version 2.0. A copy of the GPLv2 license must accompany distribution of POPSIM and work derived from it. In addition, there is ABSOLUTELY NO WARRANTY expressed or implied, and the user bears all the risk when using POPSIM. Users should be aware that POPSIM uses open-source software to perform some of its functions, including ECOS, CBC and dsOpt – copies of licenses for these products are distributed with the POPSIM package. In accepting the conditions of use on the opening screen, the user also agrees to the conditions of use for these components – please read these!

## Distribution of POPSIM

POPSIM will usually be distributed as a “zip” file. The following files are provided in the package:

- [\*popsim.xlsm\*](#) – Excel workbook used to provide simulation parameters to POPSIM and managing output of results
- [\*popsim.exe\*](#) – the executable binary for running POPSIM;
- [\*opsel.exe\*](#) – the executable binary for running OPSEL, required if using optimum selection options;
- [\*filesocp.exe\*](#) – prepares the data for the unequal contributions solver, ECOS, required if using optimum selection by Second-Order Cone Programming;
- [\*ecos-wrapper.exe\*](#) – the ECOS solver using Second-Order Cone Programming;
- [\*dsOpt.exe\*](#) – the branch-and-bound solver used by OPSEL for optimization of equal contributions by Mixed-Integer Linear Programming;
- [\*xdesign.exe\*](#) – the application for optimal mating designs for unequal contributions;
- [\*GNUv2\\_license.txt\*](#) – GNU General Public License, Version 2, applies to OPSEL and ECOS;
- [\*Eclipse\\_license.txt\*](#) – Eclipse General Public License, Version 1, applies to CBC and dsOpt;
- [\*Help401.htm\*](#) and supporting folder [\*Help401\\_files\*](#) -- system files for POPSIM help;
- [\*POPSIM\\_Userguide.pdf\*](#) – this document;
- [\*Release Notes xx-xx-20xx.pdf\*](#) – notes on changes made since the publication of the user guide – please read these!

# Installing POPSIM

The following procedure and folder locations are recommended for installing the program. File locations could be changed, once successful operation of the program has been verified, but path names should be kept very short, as per these examples.

1. Extract the zip file to a convenient folder on your machine. Some files listed above may have had the ".exe" extensions stripped, to enable the sending of the zip file through institution firewalls; if so, you will need to restore the names by adding ".exe" to distributed file names.
2. The Excel workbook file ("*popsim.xlsx*") should be put in a separate folder, such as *C:\popsim\user*. Note: this is also where POPSIM will create all of its work files during execution.
3. The "*POPSIM.exe*", "*help401.htm*" and the "*help401\_files*" folder should be put together in a folder, such as *C:\popsim\bin*.
4. The optimization applications (*opse.exe*, *dsOpt.exe*, *filesocp.exe*, *ecos.exe* and *xdesign.exe*) should be placed together in a folder. In this manual, we assume they have been placed in a separate applications folder, *C:\popsim\apps*.
5. Check the Excel settings. In Office2016 (other versions are similar, but locations may vary) click on "File" on the menu bar and select "options". For the frontend to operate correctly, please check the following settings and adjust if required:
  - Excel Options → Trust Center → Trust Center Settings ... → Macro Settings "Enable all macros"
  - Excel Options → Save → delete any default file location, if one is specified.
6. Run the example scenarios supplied with POPSIM, to verify installation. The "*popsim.xlsx*" file supplied with POPSIM is pre-loaded with examples that should run quickly on any machine. The examples include the use of optimal selection for seed orchards and auto-generation of unbalanced mating designs for unequal optimal contributions, but do not require the use of ASReml. Open the file by double clicking on the filename in File Explorer – DO NOT open the file by using the File → Open menu in Excel!! Accept the information on the splash screen, then read the information on the "Instructions" sheet, where any changes to this user guide will be noted. Then select the "MainPage" sheet. If you installed POPSIM in folders other than those suggested above, you will need to modify the folder locations for input/output and executable files. Click on "Run simulations", then "OK". The program will check the scenario data and call the main POPSIM program. Accept the information on the POPSIM splash screen. You should now see simulation progress on the simulation control page (updating iterations and cycles). If you get into trouble, terminate the program by clicking "Ctrl + Alt + Del", mark Excel in the task list and end task.

## Setting up scenarios – the input interface

POPSIM uses an Excel workbook as a user interface to help prepare control data input and manage the results output. The user will find it possible to use familiar “cut-and-paste”, formatting and file save features, as for any other Excel file. The “.xslm” file supplied with POPSIM is preloaded with example data, but this workbook can be modified by the user and saved under a new name (but always as a “.xslm” file type). Be aware that it is possible to “upset” the automated procedures built into the worksheet by adding rows or changing the macro programming. The user can always return to the example “.xslm” workbook in the distribution file to restore functionality that may be lost by inadvertent changes to the workbook.

### SET UP SCENARIOS

The process of specifying a set of simulation scenarios begins by clicking on the “**Set up scenarios**” macro button. The user is asked to specify the number of scenarios to be run, and whether you want to delete all data in the sheet beyond this (this area of the sheet is often useful for making run notes, as any columns after the final scenario will be ignored by POPSIM). The “**Set up scenarios**” macro will generate the specified number of scenarios as columns with sequential values for **KSCEN** as headings in row 4. The macro performs other set up functions in the background and **MUST** be used to generate the Scenario numbers; entering values manually for **KSCEN** will not work properly.

### SUPPLY SIMULATION CONTROL PARAMETERS

Column A on the MainPage gives an abbreviated description of each control parameter. Column B gives the short name for each parameter and indicates if the parameter is specified as an Integer (green background) or Real numeric input (orange background). POPSIM will ignore any parameter that is not required; for example, the value of **RgOP** is ignored if progeny testing uses polycross families or if no progeny testing is performed. The control data are grouped and discussed in the following sections, line-by-line, as they appear on the MainPage sheet of the “.xslm” file. Reminder notes have also been inserted as “comments” in this sheet, which serve as reminders to help select the correct options.

The control parameters are organized on the MainPage in a logical sequence and grouped by the aspect of the simulation they control, indicated by heavy borders around parameters listed in columns A and B.. The following sections follow this grouping to discuss the parameters and their use in detail.

### SCENARIO IDENTIFICATION:

#### **KSCEN** Scenario number

Each column is identified by a scenario sequence number, beginning at 1 and continuing to the number of scenarios specified by “**Set up scenarios**” macro. These numbers are auto-generated by the “**Set up Scenarios**” macro button, as described above. Do NOT try to add or modify these numbers manually.

## SIMULATION DIMENSIONS:

The time required to complete a given scenario is largely determined by how many generations (breeding cycles) are to be simulated, and how many times the simulation is to be repeated (iterations) to produce results with useful means and standard deviations.

### **NGEN** Number of generations

Specify the number of breeding cycles to be completed. The simulation will report results for the FO, F1, ..., F<sub>NGEN</sub> generations.

### **NITER** Number of iterations

Specify the number of iterations of the simulation. The simulation will be repeated **NITER** times, and the results reported as the mean outcomes and standard deviations for these iterations. Of course, larger numbers of iterations will better estimate the expectation as the central tendency of the reported outcomes, but will of course add to the execution time. **NITER** should be set to at least 100 for reliable estimates of outcome expectations, while smaller numbers can be used for approximate results or preliminary experiments.

Setting **NITER** to zero invokes a special run through a single iteration of the simulation, with a “Pause” after each cycle. This allows the user to inspect the various temporary files that are recreated after each cycle and can be valuable for understanding the outcome of selection on the structure of the various selected populations. Press “Enter” to restart the simulation after each pause.

## GENETIC STRUCTURE OF THE MANAGED POPULATION:

All strategy simulations require a description of the population genetic structure. This defines the various genetic and environmental effects and their distributions that are sampled during formation of the base founder population and generation of progeny.

### **NFndrs** Number of founders

This is the size of the original population of founders, sampled from an “infinite” base or “wild” population. For tree breeding applications, it is usually convenient to consider this to be the original pool of “plus-trees” sampled from the natural forest. It is very common that this pool will be supply a smaller number of breeding parents, selected on the basis of OP or polycross progeny testing (see below).

### **NT** Number of parents in the breeding population

Specify the size of the population of parents bred each cycle to form a recruitment population. If **NT** < **NFndrs**, the original sample of founders will first be progeny tested to identify the best **NT** parents.

### **EFFBAR(4)** Mean of trait in base population

Specify the mean of the trait (or index of traits) in the original base population of **NFndrs**.



**BID Rate of inbreeding depression**

Inbreeding depression is simulated by adjusting the dominance effect by the regression of phenotypic value on the inbreeding coefficient (Borralho 1994).

**BID** is a regression coefficient expressing the reduction in phenotypic value per unit of inbreeding, in units of phenotypic standard deviation. The regression coefficient may be calculated from experimental data by using a linear regression of inbreeding coefficients on corresponding standard deviation units of phenotypic values. For example, inbreeding depression on the mean of selfed progeny is often 5 to 30% lower than outcrossed progeny (Griffin and Cotterill 1988; Park and Fowler 1984). If the reduction in the mean of selfed progeny ( ) is one standard deviation below the phenotypic mean, the regression coefficient **BID** = -2.0. Note that **BID** is ALWAYS a negative real number, and normally in the range of 0 to -4. Note also that there is no random within-family distribution of inbreeding depression; all full-sibs will have the same inbreeding coefficient and, thus, the same inbreeding depression.

**EFFVAR(1) Additive genetic variance in the base population**

Specify the additive genetic variance of the trait (or index of traits) in the original base population of **NFndrs**.

**EFFVAR(2) Dominance genetic variance in the base population**

Specify the dominance genetic variance of the trait (or index of traits) in the original base population of **NFndrs**.

**EFFVAR(3) Epistatic genetic variance in the base population**

Specify the epistatic genetic variance of the trait (or index of traits) in the original base population of **NFndrs**.

**EFFVAR(4) Environmental variance in base/recruitment population**

Specify the environmental variance found in the base population. This environmental variance is also applied to the generation of the recruitment populations, as they are considered to be tested in a similar, undomesticated field environment.

Note that the user does not specify the trait heritability directly, as this is a function of the variances supplied, such that:

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_D^2 + \sigma_I^2 + \sigma_E^2}$$

To run scenarios at different heritability, the usual approach is to adjust the environmental variance, **EFFVAR(4)**.

**PROGENY TESTING SPECIFICATIONS:**

Not all breeding strategies use progeny testing. Some programs use progeny testing only as a first operational step to reselect the first **NT** breeding parents from a larger number initial plus-tree founders, while subsequent cycles perform selection for breeding without a progeny-testing step. Some programs may use two-step selection for every breeding cycle, so that initial candidates selected in a first step are progeny tested prior to final selection of the breeding parents based on progeny test results.

#### **PXF Progeny test option**

A progeny test may be required, either to perform selection of breeding parents from a pool of founders in the first cycle, or to support other breeding or seed orchard decisions across all cycles. The test can be performed as a “polycross”, where a controlled cross is made with a pollen mix, or by collecting “open-pollinated” (OP) seed (unknown pollen parents from the original “wild” population). The following options can be specified:

- o No progeny testing (in which case other parameters in this section are ignored)
- 1 Polycross progeny testing with assessment of PX family means
- 2 Polycross progeny testing with BLUP assessment (requires ASReml)
- 3 OP progeny testing with assessment of OP family means
- 4 OP progeny testing with BLUP assessment (requires ASReml)

Note that when BLUP assessment is performed, the progeny test observations are carried forward and become part of the multi-generation dataset when BLUP is performed in later generations.

#### **NP Number of PX pollen parents**

If PX testing is performed, a pollen mix will be generated as a random sample of **NP** parents, drawn from the original base founder population.

#### **NX Progeny test family size**

When progeny testing is performed, test families are generated with **NX** progeny per PX or OP family.

#### **RgOP Pollen parent correlation**

When progeny testing is performed with OP families, the pollen parent may be genetically correlated with the seed parent (Squillace 1974). **RgOP** describes this as phenomenon as a correlation coefficient, with a value between 0 and 1.

#### **PXENV Environmental variance for progeny testing**

Specify the environmental variance to be applied to progeny testing. This allows the user to describe an environment different from that where the original founders or recruitment populations are assessed. If progeny testing is done as a short-term nursery test, this variance might typically be less than that found in field trial sites.

## SPECIFICATIONS FOR GENERATING RECRUITMENT POPULATIONS:

A recruitment population is generated each cycle by mating those individuals selected as parents, known as the “breeding population”. POPSIM provides several different ways of generating the progeny that form the recruitment population.

### **XOPT** Crossing option to generate recruitment population

Options available are:

- 1 **NXS** controlled crosses are carried out among the BP members, according to the design grid in sheet **MDn** (see the specification of **MDn** below).
- 2 **NXS** controlled crosses are auto-generated with equal contributions from each selected parent (the mating design sheet reference **MDn** in this case is set to zero).
- 3 Breeding is carried out with one or more polymixes, made up as subsets of parents, with or without paternity determination for a pre-selected portion of the progeny. **NT\*PMXGF** candidates will pre-selected, genotyped, reanalysed by BLUP before the final **NT** selections are made. Note that when **PMXGF** = 0, selection of the final **NT** parents is made directly without paternity determination.
- 4 As for option 3, but the phenotypic data will NOT be reanalysed by BLUP after paternal genotype is determined. When **PMXGF** = 0, options 3 and 4 are equivalent.
- 5 Use XDesign to generate a mating design for NXS crosses from optimal, unequal contributions from genotypes selected by OPSEL.
- 8 Progeny test families are generated by open pollination, in which case paternal genotypes are randomly sampled from the distributions of base population genetic effects, with pollen:seed parent correlation the same as specified in **RgOP**.

### **ISORT** Mate assortment option

Mates can be allocated at random, or they may be ranked and mated assortatively (positive assortative mating, or “PAM”). Options available are:

- 0 Random assortment of mates
- 1 Assortment on progeny-test family means (GCA values)
- 3 Assortment on phenotypic value (of seedling or mean of all ramets of a clone)
- 5 Assortment on the combined index of family mean and individual phenotype, weighted by their respective heritabilities (Baker 1986; Falconer and Mackay 1996)
- 7 Assortment on BLUP breeding values (requires ASReml)

### **NPMX** Number of pollen mixes

This is set to an integer value, describing the number of pollen mixes to be assembled for PMX breeding. **NPMX** has no effect unless **XOPT** = 3 or 4.

### PMXSIZE

When PMX breeding is specified (i.e., **XOPT** = 3 or 4), describes the number of parents used to form each pollen mix. The parents are assigned in rank order (as determined by the mate assortment option **ISORT**) to each pollen mix, and each parent can only be used once. As there are only **NT** pollen parents to choose from, the value of **NPMX\*PMXSIZE** must be  $\leq$  **NT** (note that it is not necessary to assign all **NT** parents to a pollen mix).

### PMXGF

When PMX breeding is specified (i.e., **XOPT** = 3 or 4), declares the number of individuals pre-selected for paternity determination, for every final selection to the next breeding population. For example, if **NT**=48 and **PMXGF**=10, the preselection step will identify the best 480 trees, by mass selection on phenotype or by BLUP BV. POPSIM then reveals the true paternity of each selection, simulating laboratory marker determination of paternity by exclusion. This paternity is assumed to be 100% accurate, with no possibility of contamination from sources outside the pollen mix. If **XOPT**=3, the data are then re-evaluated by BLUP with the more complete pedigree. The final selection is then made by Group-merit selection (GMS), by restricted mass selection, or by optimal selection with equal contributions (dsOpt in OPSEL). If **PMXGF** is set to zero, the preselection and paternity determination steps are skipped, and selection of the breeding population is performed directly without any knowledge of paternity.

### MaxC Relative avoidance option

Assortment may lead to mating of close relatives, so it is often desirable to avoid this by making minor substitutions after assortment.

When **XOPT** = 1, **MDn** = 0 and **NXS** specified for balanced single-pair, doublepair, or triple-pair mating (see description of **NXS** below), options for avoidance of relatives during auto-generation of the mating design have the following Integer values:

- 0 No attempt to avoid mating close relatives
- 1 Avoid mating half-sib relatives
- 2 Avoid mating full-sib relatives

When **XOPT** = 5 (mating design generated by XDesign for optimum, unequal contributions), the user can specify a maximum threshold level of coancestry tolerated between crossed parents as a Real number, where  $0.0 \leq \text{MaxC} \leq 1.0$ . XDesign will modify the provisional list of ranked matings by swapping parents, attempting to comply with the coancestry threshold. Note that this constraint may be violated if no combination of mates satisfying the constraint is available (see documentation for XDesign by Mullin 2017b).

### MDn Mating design grid reference (see Appendix 1)

Used only for crossing option **XOPT** = 1, requiring generation of the recruitment population by controlled pollination according to a mating design grid specified in the Excel sheet corresponding to the value specified **MDn**, i.e., worksheets labelled MD1, MD2, ... MD10, and the sheet number n (1, 2, ... 10)<sup>1</sup>. Alternatively, balanced single-pair, double-pair or triple-pair matings can be specified by setting **MDn** to 0, and declaring **NXS** equal to **NT**, **2\*NT**, or **3\*NT**, respectively.

<sup>1</sup> Details on how to use the mating design grid sheets in the Excel workbook can be found in Appendix I.

**NXS Number of controlled crosses**

When **XOPT** = 1 and  $1 \leq \mathbf{MDn} \leq 10$ , this parameter is auto-filled at execution time with the total number of crosses specified in the mating design grid **MDn**.

When **XOPT** = 2 and **MDn** = 0, a balanced mating design will be auto-generated, but it will be necessary here to specify the total number of crosses among the breeding parents as **NT/2**, **NT** or  $1.5 * \mathbf{NT}$ , respectively, for single-pair, double-pair, or triple-pair matings, as described under **MDn**.

When **XOPT** = 3 or 4, the **NXS** controlled crosses are with a polymix, and must equal **NT\*NPMX**.

When **XOPT** = 5, **NXS** specifies the number of crosses to be generated by XDesign, using the **NXS**\*2 parental contributions from the OPSEL solution for optimal selection with unequal contributions.

**NC Number of genotypes tested per cross**

This is the number of genotypes generated for a controlled-cross family in the recruitment population, or “family size”.

**NR Number of ramets tested per genotype**

This parameter describes the number of clonal replicates tested per genotype. If the recruitment population is propagated as seedlings, **NC** should be declared as 1.

**SELECTION OF THE BREEDING POPULATION:**

Once a recruitment population has been generated, the next step in an improvement strategy will be selection of a parents to form the breeding population for the next cycle. Again, POPSIM provides several options for how this is done.

**IBPSEL Selection method for BP selection**

Numerous options are available for selection of breeding parent candidates from the recruitment population. These include standard approaches such as mass selection and combined-index (CI) selection (Baker 1986; Falconer and Mackay 1996), as well as combining gain and diversity targets with “Group-merit” selection (GMS) (Lindgren and Mullin 1997), optimum selection using the Mixed-Integer Linear Programming option for equal contributions in OPSEL (Mullin and Belotti 2016), or the Second-Order Cone Programming option for unequal contributions in OPSEL (Yamashita et al. 2018).

---

<sup>2</sup> The user is referred to Appendices II and III for further information on combined-index selection and group-merit selection, respectively.



The specific options available are:

- 0 Random (no selection is applied in the random sampling of **NT** parents)
- 1 Mass selection (selection is based on individual phenotypic value or clone mean)
- 3 CI selection with weights on family and individual values corresponding to their respective heritabilities, calculated from the test data
- 4 CI selection with weights defined by a user-specified individual heritability given under **WEIGHT** (see below)
- 5 GMS on phenotype
- 6 GMS on combined index
- 7 GMS on BLUP breeding value (requires ASReml)
- 8 Mass selection on BLUP breeding value (requires ASReml)
- 9 Optimum selection on phenotype with equal contributions – using the Mixed-Integer Linear Programming solver dsOpt in OPSEL.
- 10 Optimum selection on BLUP breeding value with equal contributions – using the Mixed-Integer Linear Programming solver dsOpt in OPSEL (requires ASReml)
- 11 Optimum selection on phenotype with unequal contributions – using the Second-Order Cone Programming solver ECOS in OPSEL.
- 12 Optimum selection on BLUP breeding value with unequal contributions – using the Second-Order Cone Programming solver ECOS in OPSEL (requires ASReml)

**LGENBP Breeding population selection across generations?**

All selection options in POPSIM (**IBPSEL**) permit the user to select the next breeding population from among the progeny of the previous cycle breeding population, by setting **LGENBP** = 1. With this option, generations remain discrete, with no overlap from one generation to the next.

When **IBPSEL** = 11 or 12 (optimum selection with unequal contributions) we can also set **LGENBP** = 2 so that selection considers the progeny of the current cycle as candidates, as well as ALL of their ancestors, regardless of generation. We assume that the ancestor genotypes are maintained in an archive and are still available for breeding.

**WEIGHT User-specified individual heritability (CI selection), weight on group coancestry (GMS), or constraint on coancestry for optimum selection**

This parameter is a real number used supply a critical variable for different types of selection.

If **IBPSEL** has the value of 4 (combined-index selection with specified weight), the **WEIGHT** is a user-provided individual heritability, which is used in the calculation of index weights, rather than using the individual heritability estimated from the test data.

If **IBPSEL** has a value of 5, 6 or 7, this **WEIGHT** is applied to the group coancestry in the calculation of group-merit values.

If **IBPSEL** has a value of 9, 10, 11 or 12, the **WEIGHT** parameter is the constraint on group coancestry passed to **OPSEL** and applied during optimum selection.

For all other settings of **IBPSEL**, any value specified for **WEIGHT** is ignored.

**NBPFS Maximum number of selections from a full-sib family**

This parameter is used to apply restrictions to some breeding population selection methods on the maximum number of progeny from a full-sib family that can be selected. The manner in which the restriction applies varies slightly among selection methods as follows:

When **IBPSEL** = 0, 1, 3, 4 or 8, **NBPFS** is applied as a limited to how many individuals can be selected from the same full-sib family. If the user sets **NBPFS** = 32000, the restriction on number of full sibs is NOT applied. If **NBPFS**  $\neq$  32000, it must also be  $\leq$  **NC** (i.e., no larger than the family size of a cross).

When **IBPSEL** = 5, 6 or 7 (i.e., selection is on Group Merit with a weight on group coancestry), the value of **NBPFS** is ignored.

When **IBPSEL** = 9 or 10 (i.e., selection is optimized for equal contributions with a constraint on group coancestry), the value of **NBPFS** is used to truncate the candidate list of current cycle progeny to include only the best **NBPFS** individuals from each family.

When **IBPSEL** = 11 or 12 (i.e., selection is optimized for UNEqual contributions with a constraint on group coancestry), the value of **NBPFS** is used to declare the maximum contribution of a given genotype to the selection solution.

**NBPHS Maximum number of half-sib relatives**

This parameter is used only when **IBSEL** = 0, 1, 3, 4 or 8, in which case **POPSIM** will restrict selection from the recruitment population such that no more than **NBPHS** previous selections are related to the selection candidate as half-sib relatives. In this case, if the user sets **NBPHS** = 32000, the restriction on the number of half sibs is NOT applied, but if **NBPHS**  $\neq$  32000, it must also be  $\leq$  **NC** (i.e., no larger than the family size of a cross).

For all other settings of **IBPSEL**, this parameter is ignored.

**NBPPC Maximum number of progeny per parent**

This parameter is used only when **IBSEL** = 0, 1, 3, 4 or 8, in which case **POPSIM** will restrict selection from the recruitment population such that any given parent of a selection candidate does not contribute any more than **NBPPC** selections. In this case, if the user sets **NBPCC** = 32000, the restriction on the number of progeny per parent is NOT applied.

For all other settings of **IBPSEL**, this parameter is ignored.

### NBPPMX

This parameter only applies to polycross breeding situations, i.e., where **XOPT** = 3 or 4. Selection of candidates for paternity determination is limited to the best **NBPPMX** candidates per PMX family (mass selection on either phenotype or BLUP BV, depending on **IBPSEL**). The final selection is performed using the method declared in **IBPSEL**, but limited to selection among preselected candidates whose paternity has been determined.

### PSType Preselection type

Some breeding strategies require a form of pre-selection for one or more cycles. The following options are available:

- 0 No preselection imposed. The selection of **NT** candidates is made directly from the recruitment population for all cycles
- 1 Preselection of **NT\*PSF** candidates from the recruitment population is followed by estimation of candidate GCAs through progeny testing, in order to impose a final selection of **NT** breeding parents from among the preselected candidates. This is equivalent to the former selection-followed-by-progeny-testing strategy used in Sweden.
- 2 This is a special option that allows **NT/PSF** initial founders to give rise to a breeding population of size **NT** after the first cycle. In other words, the F1 recruitment population arises from **NT/PSF** founders, while the F2 and subsequent recruitment populations arise from a larger breeding population of **NT** parents. This is a special option representing “Strategy 5” for the Scots pine programme in Sweden.

### PreSelCyc Preselection cycles

When **PSType** has been declared with a value of 1, the two-step preselection followed by progeny testing leading to final selection of the breeding population, is carried out for **PreSelCyc** cycles. After this number of cycles, POPSIM reverts back to **PSType** = 0 for all remaining cycles.

### PSF Selection factor

This parameter is used, as describe above under **PSType**, to determine the number of candidates per final selection to be progeny tested (**PSType**=1), or the expansion of original **NT/PSF** founders to form a breeding population of **NT** parents (**PSType**=2).

## OPTIMUM SELECTION SWITCHES

The following three parameters are used only for BP selection methods **IBPSEL** = 9 or 10, which use a branch-and-bound solver to optimize selection of **NT** genotypes that will contribute equally to the next recruitment population. The algorithm is described in Mullin and Belotti (2016). The switches are ignored for all other selection methods.

### SWITCH-t Maximum CPU seconds per selection cycle

The branch-and-bound algorithm will generally find a very good solution rather quickly, but it may take some time to confirm that the solution is truly the best possible. The solver will continue until the current solution is confirmed as the true optimum. This parameter limits the search time to the specified number of CPU seconds.

**SWITCH-n Maximum number branch nodes per selection cycle**

This parameter provides another way to limit continued execution of the optimum solver by specifying the maximum number of nodes explored by the branch-and-bound algorithm. Few users will find this useful, as **SWITCH-t** and **SWITCH-g** provide more intuitive ways to limit the search.

**SWITCH-g Threshold gap % between current and best possible**

This parameter causes the solver to terminate once it has confirmed a solution within the specified % from the theoretical best. While it may be possible to improve the solution, the user may feel that beyond this point there is little return from the additional execution time (which will, of course, ultimately end when **SWITCH-t** is triggered).

## **SELECTION OF SEED ORCHARD PRODUCTION POPULATION:**

While not required, most users will want to evaluate progress based on changes in a “production population”. A seed orchard is perhaps the most common way of producing improved material to be established in operational forest plantations. All individuals selected into a breeding population are considered to have been archived and available for grafting into an open-pollinated seed orchard. POPSIM will report statistics for individuals selected into the orchard for each cycle, including the genetic gain expected under “ideal” orchard conditions, i.e., random panmictic mating, equal numbers of gametes per parent, and no impact of contaminating pollen from outside the orchard.

**NSSO Number of selections for seed orchards**

This parameter applies ONLY for seed orchard selection methods **ISOSEL** = 1, 2, 3 or 4, and determines how many genotypes will be selected in each cycle to form a seed orchard production population, where each genotype is represented in the orchard by an equal number of ramets.

**ISOSEL Selection method advancing archived parents to seed orchards**

Selection to the orchard need not be based on the same criteria as applied to select the breeding population from the recruitment population, but only those individuals selected previously as members of a breeding population are eligible to advance to the orchard as these are likely to be represented in archives and available for orchard establishment. The following options are available:

- 1 Forward selection on phenotype (or clone mean), i.e., without any additional progeny-test estimation of GCA
- 2 Forward combined-index selection without any additional progeny-test estimation of GCA
- 3 Forward selection on BLUP breeding value (estimated without additional progeny testing)
- 4 Backward selection on progeny-test GCA or BV estimates (as defined under **PXF** – see “Progeny testing specifications” above)
- 5 OPSEL called to optimize orchard selection with unequal numbers of ramets per clone, with the same constraint on group coancestry applied each cycle, determined by the value of theta.
- 6 As for 5, but the constraint on group coancestry is increased each cycle by the value of theta.

#### **LGEM Selection across generations?**

Candidates available for seed orchard establishment may be limited to those parents in the current cycle, or they may include breeding population members from previous cycles that have been maintained in archives. The options available are:

- 1 Orchard selections originate only from current cycle (available for all options for **ISOSEL**)
- 2 Orchard selections can originate from the current or any previous cycle (not available for **ISOSEL** = 1 or 2, but available for all others)

#### **Theta Constraint on seed orchard group coancestry**

Used only for seed orchard optimum selection options **ISOSEL** = 5 or 6, **Theta** is a real number describing the group coancestry constraint applied during selection. Note that the Status Number of the orchard will be  $\geq 0.5/\text{Theta}$ .

### **ORCHARD SELECTION RESTRICTIONS**

For orchard selection methods that do NOT use the optimization algorithm, i.e., **ISOSEL** = 1, 2, 3, or 4, restrictions can be imposed on the number of orchard selections made from a given cross, the number of half-sib relatives for a given selection, and/or the number of selections related to a given parent: **NSOFS**, **NSOHS** and **NSOPC**, respectively. A value for any of these restrictions of 32000 has the effect of ignoring the restriction. Otherwise, POPSIM will attempt to satisfy all other restrictions with values set at less than 32000. The user is warned that this can lead to unexpected results, as POPSIM may be forced to “relax” some restrictions to find selections that satisfy all restrictions.

#### **NSOFS Maximum number of orchard selections from a full-sib family**

**NSOFS** declares the maximum number of orchard selections that can come from a given full-sib family.

#### **NSOHS Maximum number of half-sib relatives**

POPSIM will restrict selection to the orchard such that no more than **NSOHS** previous selections are related to a selection candidate as half-sib relatives.

#### **NSOPC Maximum number of progeny per parent**

POPSIM will restrict selection to the orchard such that any given parent does not contribute any more than **NSOPC** selections.

### **SELECTION OF A FULL-SIB FAMILY PRODUCTION POPULATION:**

Tested families can also be used as a production population, either by repeating the cross with conventional controlled pollination or by mass control-pollinated orchards, and/or vegetative multiplication of seed (Bramlett 1997; Bridgwater et al. 1998; Carson et al. 1992; Harbard et al. 1999; Sweet et al. 1992), commonly referred to as “family forestry” (Libby 1992). POPSIM can assemble such a production population from the crosses tested in the current recruitment population.

#### **NFAM Number of families selected for deployment**

This is the number of families selected for deployment among those tested in the current-cycle recruitment population. If a FS family production population is not required, set **NFAM** = 0.



**NFAMHS Number of half-sib related families**

Once a family has been selected to the mixture, **NFAMHS** restricts how many families related at the half-sib level can be added to the mix.

**NFAMPC Number of families with a common parent**

POPSIM will restrict selection such that any given parent does not contribute any more than **NFAMPC** families selected for deployment.

**SELECTION OF A MIXTURE OF SELECTED CLONES:**

When the recruitment population has been deployed with clonal replicates, there is an opportunity to select the best clones and deploy these as a mixture. Regeneration of clones is problematic for many forest trees, as physiological ageing limits the feasibility of propagating stored material, unless the cloning has been by somatic embryogenesis and the cultures can be retrieved from cryo-storage (Park 2002). POPSIM can assemble such a production population from the genotypes clonally replicated in the current recruitment population.

**NCLN Number of clones selected for deployment**

This is the number of clones selected for deployment among those genotypes clonally replicated in the current-cycle recruitment population. If a clonal production population is not required, set **NCLN** = 0.

**ICLNSEL Method for selecting clones for the deployed mixture**

Selection to the deployed clone mix need not be based on the same criteria as applied to select the breeding population from the recruitment population, but only clones from the current-cycle's recruitment population are eligible. The following options are available:

- 1 Forward selection based on clone mean
- 2 Forward selection on combined index value
- 3 Forward selection on BLUP breeding value

**NCLNFS Maximum number of clones selected from a full-sib family**

**NCLNFS** declares the maximum number of clones that can be selected from a given full-sib family.

**NCLNHS Maximum number of half-sib relatives**

In order for POPSIM to consider a clone for selection to the mixture, there must be no more than **NCLNHS** previous selections that are related to the candidate as half-sibs.

**NCLNPC Maximum number of progeny per parent**

POPSIM will restrict selection of clones such that any given parent does not contribute any more than **NCLNPC** selections.

**LOCAL SETTINGS:**

Folder locations specified in the Excel worksheet should have path names that contain NO blank characters, as these may cause issues with some options. The examples given on the worksheet are known to work well on all machines.

The folder for scenario input/output is the one containing the Excel workbook (example: C:\Popsim\User\). This is also where POPSIM will manage its various work files. It is not necessary to specify this folder on the *.xls*m, but its path name should be kept short.

The following additional folders must be specified for ANY POPSIM run, as these define locations where critical files are found:

**Folder containing popsim.exe:** It is usually convenient to store POPSIM and its Help files in a separate folder, as this allows them to be updated easily (example: *C:\Popsim\bin\*).

Folders must also be specified if one or more of the scenarios uses either ASReml (for BLUP estimation of breeding values), or OPSEL (for optimal selection of breeding or seed orchard populations), or XDesign (for generation of mating designs), as follows:

**Folder containing ASReml.exe:** this location is defined during the installation of ASReml (example *c:\Program Files\Asreml4\Bin\*).

**Folder containing optimal selection and/or mating apps:** If optimal selection or mating is required, these apps and their supporting files must be stored in a common apps folder (example *c:\Popsim\Apps\*). Remember to keep the path name short!

## Running a series of scenarios

Once the scenarios have been described on the Excel workbook MainPage, be sure to save as a macro-enabled workbook – a unique name can be assigned. To run POPSIM, the user **MUST** open the workbook from its folder location by double clicking on the file name – do NOT use the Excel Open → File command or attempt to open from a “Recent files” listing in File Explorer!

From the MainPage sheet, click on the “**Run Simulations**” macro button to launch the simulation. The user will be asked to specify which scenario to start the simulation – usually this would be “1”. The macro will conduct a general check of the data in the worksheet, to ensure that a readable control file called “*input.txt*” can be generated. The macro then calls “*popsim.exe*”, which is the main simulation engine. POPSIM will open a splash screen window which gives legal information, and also reports the “Build date” of the program. This build date should be given in any correspondence requesting support. This window can be resized to allow other windows to be visible on the desktop, while the simulation is running. The Excel workbook can also be closed, as POPSIM does not require it to complete the simulation request.

POPSIM will conduct a thorough check of all control data for all scenarios. Any warnings or errors will be displayed, to guide in the correcting of scenario data. If the control data are error-free, the program will begin the simulation of the first scenario. The program provides an on-screen progress report, stating what step is being executed, in what iteration, generation and scenario, and also gives the total execution time since the program was launched.

Execution time can vary tremendously, depending on the dimensions, number of iterations requested, complexity and analysis procedures required. If BLUP calculations are required, the simulation will periodically make calls to ASReml, which will launch in a separate window. Similarly, if optimum selection is required, POPSIM will make calls, as required, to OPSEL and possibly to XDesign, which will execute in separate windows.

It is not unusual for scenarios to take several minutes per iteration, and several scenarios may require several hours, sometimes even days, to complete. If at all possible, avoid using the computer for other tasks while POPSIM is running. Multi-tasking response time will be poor, and you could inadvertently interrupt the simulation.

Once POPSIM has completed all iterations of all scenarios, the program will sound a “beep” and display a message on the screen that the “Program performed all tasks”. Hit ENTER to terminate the program and close the window. Return to the Excel interface workbook to retrieve and view your results, as described below.

## Retrieving and viewing output

POPSIM results are stored on completion of each scenario in a file called “*output.txt*”, but are best viewed through the Excel interface. Open the Excel macro-enabled control workbook used to launch the simulation, by double-clicking on the file name and accept the splash screen. Select the “Results” sheet, and click on the “**Read Output**” macro button to load the results.

POPSIM will read one line of results for each generation of each scenario. The results are arranged in columns describing the mean and standard deviation over all iterations, for each population parameter. The breeding population is listed first. Scroll to the right and you will find comparable data for the seed orchard and other production populations that may have been requested.

## Saving scenario parameters and results

After completing a series of scenarios and loading the results into the Excel control data workbook, it is usually desirable to save a permanent record of the results and the control parameters that generated them. On the “Results” sheet, click on the “**Save Results**” macro button. Declare a suitable location and file name.

The resulting workbook contains 3 worksheets. The control data are stored in the first “Simulation Parameters” sheet, the results in the “Results” sheet, and a third “User Page” is provided as a place to conduct analyses of the results.

## Analysing simulation results

Different population parameters will be of interest to different users, depending on the objective of their simulation. The usual analysis procedure involves graphing the development of different population parameters against time. The simulation results are ordered by generation, but the user may have different timelines in mind for different scenarios. Graphs can be prepared directly in the Results workbook, using the various tools provided in Excel. Macros can also be easily written in Excel to retrieve the parameters of interest and display these in custom formats or graphs, if many scenarios are to be analysed

## When things go wrong!

The user should be aware that POPSIM is in a continuing state of development; new features are added, and others modified, in response to user-feedback and the needs of new research. Every user will ask different things of the simulation, and there will be times when development testing has not covered all possibilities. Unexpected things can happen!

POPSIM may provide messages from the Excel control data workbook or from the main program. Quite often, messages are generated when the user is attempting to declare an “illegal” combination of management parameters. Reading the messages, combined with the parameter descriptions given in this manual, may help to identify the cause of the problem.

POPSIM has a very large combination of parameters that can be controlled by the user, and not all combinations have been thoroughly tested during development. If POPSIM accepts all of the initial data as legal, there may still be times when the program crashes, with or without a message. If you cannot rectify a problem, you can request support by writing to: [tim.mullin@skogforsk.se](mailto:tim.mullin@skogforsk.se)

When asking for help, it is very important to provide the following to help reproduce and diagnose the problem:

1. Copy of the Excel control data workbook with the scenario(s) that caused the problem.
2. A full description and screen shot of any messages produced (the key combination Ctrl+PrtScn will put a screen shot in the clipboard, and this can then be pasted into a Word document or body of an email).
3. The “Build date” of the POPSIM main program (given on the main splash screen when the simulation starts), or the file date of your “popsim.exe” file.
4. Versions of ASReml, OPSEL and XDesign (if used).

## Help to improve POPSIM

POPSIM has developed with many ideas and contributions of tree breeders and geneticists in many countries. Users are encouraged to become engaged, to ask questions and to make suggestions for future capabilities. Users are encouraged to subscribe to the email updates at [www.skogforsk.se/popsim](http://www.skogforsk.se/popsim).

## References

- Andersson, E.W., Spanos, K.A., Mullin, T.J. & Lindgren, D. 1998. Phenotypic selection can be better than selection for breeding value. *Scand. J. For. Res.* 13(1): 7-11.
- Baker, R.J. 1986. Selection indices in plant breeding. CRC Press, Boca Raton, FL.
- Borralho, N.M.G. 1994. Heterogeneous selfing rates and dominance effects in estimating heritabilities from open-pollinated progeny. *Can. J. For. Res.* 24(5): 1079-1082.
- Bouffier, L., Raffin, A & Mullin, T.J. 2011. Evaluation of polymix breeding with paternity analysis in the maritime breeding program. In *Noveltree Third Annual Meeting*, 4-8 April 2011, Madrid, Spain.
- Bramlett, D.L. 1997. Genetic gain from mass controlled pollination and topworking. *J. For.* 95(3): 15-19.
- Bridgwater, F.E., Bramlett, D.L., Byram, T.D. & Lowe, W.J. 1998. Controlled mass pollination in loblolly pine to increase genetic gain. *For. Chron.* 74(2): 185-189.
- Carson, M.J., Vincent, T.G. & Firth, A. 1992. Control-pollinated and meadow seed orchards of radiata pine. In *Proceedings of IUFRO-AFOCEL Symposium on Mass Production Technology for Genetically Improved Fast Growing Forest Tree Species*, Bordeaux, France, 14-18 September 1992. AFOCEL, Nangis, France. pp. 100-109.
- Cockerham, C.C. 1967. Group inbreeding and coancestry. *Genetics*, 56: 89-104.
- Falconer, D.S. & Mackay, T.F.C. 1996. Introduction to quantitative genetics. 4th ed. Prentice Hall, Harlow, UK.
- Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52: 399-433.
- Fraser, A.S. 1957. Simulation of genetic systems by automatic digital computers. I. Introduction. *Aust. J. Biol. Sci.* 19: 484-491.
- Gilmour, A.R., Gogel, B.J., Cullis, B.R. & Thompson, R. 2009. ASReml User Guide, Release 3.0. VSN International Ltd., Hemel Hempstead, UK.
- Griffin, A.R. & Cotterill, P.P. 1988. Genetic variation in growth of outcrossed, selfed and open-pollinated progenies of *Eucalyptus regnans* and some implications for breeding strategy. *Silvae Genet.* 37(3-4): 124-131.
- Harbard, J.L., Griffin, A.R. & Espejo, J. 1999. Mass controlled pollination of *Eucalyptus globulus*: a practical reality. *Can. J. For. Res.* 29(10): 1457-1463.
- Libby, W.J. 1992. Use of genetic variation for breeding forest trees. In *Plant breeding in the 1990s*. CAB International, UK, Wallingford. pp. 101-117.
- Lindgren, D., Danusevičius, D. & Rosvall, O. 2008. Balanced forest tree improvement can be enhanced by selecting among many parents but maintaining balance among grandparents. *Can. J. For. Res.* 38(11): 2797-2803.
- Lindgren, D., Gea, L. & Jefferson, P. 1996. Loss of genetic diversity monitored by status number. *Silvae Genet.* 45(1): 52-59.
- Lindgren, D. & Mullin, T.J. 1997. Balancing gain and relatedness in selection. *Silvae Genet.* 46(2-3): 124-129.



- Lstibůrek, M., Mullin, T.J. & El-Kassaby, Y.A. 2006. The impact of differential success of somatic embryogenesis on the outcome of clonal forestry programs. I. Initial comparison under multitrait selection. *Can. J. For. Res.* 36(6): 1376-1384.
- Lstibůrek, M., Mullin, T.J., Lindgren, D. & Rosvall, O. 2004a. Open-nucleus breeding strategies compared with population-wide positive assortative mating. I. Equal distribution of testing effort. *Theor. Appl. Genet.* 109(6): 1196-1203.
- Lstibůrek, M., Mullin, T.J., Lindgren, D. & Rosvall, O. 2004b. Open-nucleus breeding strategies compared with population-wide positive assortative mating. II. Unequal distribution of testing effort. *Theor. Appl. Genet.* 109(6): 1169-1177.
- Lstibůrek, M., Mullin, T.J., Mackay, T.F.C., Huber, D. & Li, B. 2005. Positive assortative mating with family size as a function of predicted parental breeding values. *Genetics*, 171(3): 1311-1320. doi: 10.1534/genetics.105.041723.
- Mullin, T.J. 1994. Evaluating the economics of alternative breeding and deployment strategies for northeastern conifers. In *Proceedings of the 24th Meeting of the Canadian Tree Improvement Association, Part 2, 9-12 August 1993, Fredericton, NB.* Edited by J. Lavereau. Natural Resources Canada, Fredericton, NB. pp. 82-105.
- Mullin, T.J. 2017a. OPSEL 2.0: a computer program for optimal selection in tree breeding. Arbetsrapport från Skogforsk Nr 954-2017. Skogforsk, Uppsala, SE.
- Mullin, T.J. 2017b. XDesign 1.0: a tool for generation of complex mating designs with optimal contributions. Arbetsrapport från Skogforsk Nr 956-2017. Skogforsk, Uppsala, SE.
- Mullin, T.J. & Belotti, P. 2016. Using branch-and-bound algorithms to optimize selection of a fixed-size breeding population under a relatedness constraint. *Tree Genet. Genom.* 12(4).
- Mullin, T.J., Hallander, J., Rosvall, O. & Andersson, B. 2010. Using Simulation to Optimise Tree Breeding Programmes in Europe: An Introduction to POPSIM. Arbetsrapport från Skogforsk Nr 711, Skogforsk, Uppsala, SE.
- Mullin, T.J. & Park, Y.S. 1995. Stochastic simulation of population management strategies for tree breeding: a new decision-support tool for personal computers. *Silvae Genet.* 44(2-3): 132-141.
- Park, Y.-S. 2002. Implementation of conifer somatic embryogenesis in clonal forestry: technical requirements and deployment considerations. *Annals of Forest Science* 59(5/6): 651-656.
- Park, Y.S. & Fowler, D.P. 1984. Inbreeding in black spruce (*Picea mariana* (Mill.) B.S.P.): self fertility, genetic load, and performance. *Can. J. For. Res.* 14(1): 17-21.
- Rosvall, O., Kroon, J. & Mullin, T.J. 2015. Optimized breeding strategies at equivalent levels of population diversity. Arbetsrapport från Skogforsk Nr. 882-2015.
- Rosvall, O., Lindgren, D. & Mullin, T.J. 1998. Sustainability, robustness and efficiency of a multi-generation breeding strategy based on within family clonal selection. *Silvae Genet.* 47(5-6): 307-321.

- Rosvall, O. & Mullin, T.J. 2003. Positive assortative mating with selection restrictions on group coancestry enhances gain while conserving genetic diversity in long-term forest tree breeding. *Theor. Appl. Genet.* 107(4): 629-642.
- Rosvall, O. & Mullin, T.J. 2013. Introduction to breeding strategies and evaluation of alternatives. In *Best Practice for Tree Breeding in Europe*. Edited by T.J. Mullin and S.J. Lee. pp. 7-27.
- Rosvall, O., Mullin, T.J. & Lindgren, D. 2003. Controlling parent contributions during positive assortative mating and selection increases gain in long-term forest tree breeding. *For. Genet.* 10(1): 35-53.
- Shelbourne, C.J.A. 1969. Tree breeding methods - A review of tree breeding strategies in relation to classical plant breeding methods with quantitative genetic expectations of gain, and predictions using data from *Pinus radiata*. New Zealand Forest Service, Wellington.
- Shelbourne, C.J.A. 1992. Genetic gains from different kinds of breeding population and seed or plant production population. Paper presented at the IUFRO symposium "Intensive forestry: the role of eucalypts", held in Durban, South Africa, in September 1991. *South African Forestry Journal* 160: 49-65.
- Squillace, A.E. 1974. Average genetic correlations among offspring from open-pollinated forest trees. *Silvae Genet.* 23(5): 149-156.
- Sweet, G.B., Dickson, R.L., Donaldson, B.D. & Litchwark, H. 1992. Controlled pollination without isolation - a new approach to the management of radiata pine seed orchards. *Silvae Genet.* 41(2): 95-99.
- Verry, S.D., Snedden, C.L. & Parfitt, R.C. 2000. Program for the deterministic modelling of genetic gains of tree breeding and seed and clone production strategies. *Southern African Forestry Journal*(189): 3-9.
- Weng, Y.H., Park, Y.S., Simpson, D. & Mullin, T.J. 2009. Efficiencies of clonally replicated and seedling testing for spruce breeding and deployment strategies. *Silvae Genet.* 58(5/6): 292-299.
- White, T.L., Adams, W.T. & Neale, D.B. 2007. *Forest genetics*. CABI Publishing, Wallingford, Oxfordshire, UK.
- Yamashita, M., Mullin, T.J. & Safarina, S. 2018. An efficient second-order cone programming approach for optimal selection in tree breeding. *Optim Lett.* doi: 10.1007/s11590-018-1229-y.

## APPENDIX I

### How to use the mating design function in the POPSIM Excel workbook

POPSIM provides various methods to generate a recruitment population. The crossing option **XOPT** = 1 invokes controlled crossing among parents in the breeding population, using a mating design grid specified by **MD**. Specifying **MDn** = 1, instructs POPSIM to use the mating design described on the **MD1** worksheet.

To format the design grid, the user clicks on the “**Set up Parents**” button, and enters the number of trees in the breeding population – this should equal **NT**. The worksheet will clear all cells and draw a red border around an **NT** x **NT** grid. If the mating is to be positive assortative, parent 1 is the highest ranked and the **NT**th parent the lowest.

To instruct POPSIM to perform a cross, the user places an “x” in the appropriate cell. The figure illustrates a mating design to make 6 crosses among 6 parents, i.e., double-pair mating.

	A	B	C	D	E	F	G	H	I	J	K	L	M
2													
3													
4													
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													
22													

Set up Parents

MAT

No. of parents (per subline)

6

Mark each cross by

x

	1	2	3	4	5	6							
1		x	x										
2				x									
3					x								
4						x							
5							x						
6								x					

## APPENDIX II

### How POPSIM calculates “combined index”

"Combined-index selection" is based on an individual index value ( $CI$ ) that weights family and individual performance by their respective heritabilities (Falconer and MacKay 1996):

$$CI = h_w^2 (\bar{X}_{ifm} - \bar{X}_{fm}) + h_F^2 (\bar{X}_{fm} - \bar{X})$$

where  $\bar{X}_{ifm}$  is the mean (or individual observation) for the  $i$ th genotype in the  $fm$ th family;  $\bar{X}_{fm}$  is the mean performance of the  $fm$ th full-sib family of size  $n$ ;  $\bar{X}$  is the overall mean;  $h_F$  is the heritability of full-sib family means, and calculated as

$$h_F^2 = h^2 \frac{1 + (n-1)r}{1 + (n-1)t} ;$$

$h_w^2$  is the heritability of within-family deviations, calculated as

$$h_w^2 = h^2 \frac{(1-r)}{(1-t)} ;$$

$h^2$  is the narrow-sense heritability of individual observations;  $r$  is the genotypic intraclass correlation, which is  $1/2$  for full-sib families;  $t$  is the phenotypic intraclass correlation;  $n$  and is the family size.

Baker (1986) showed that for families of large size,  $t = rh^2$ , so in the case of full-sib families the relevant heritability formulae become:

$$h_F^2 = h^2 \frac{1 + \frac{1}{2}(n-1)}{1 + \frac{1}{2}(n-1)h^2} \quad \text{and} \quad h_w^2 = h^2 \frac{\frac{1}{2}}{1 - \frac{1}{2}h^2}$$

## APPENDIX III

### How POPSIM calculates “group merit”

Group-merit selection is a technique first described by Lindgren and Mullin (1997), where the objective is to maximise gain while constraining relatedness so that genetic diversity is conserved. The approach is to define a selection criterion for a set of genotypes that combines both breeding value and diversity, and then to apply a procedure that searches for an optimal trade-off between them.

For the purpose of calculating group merit, breeding value ( $g$ ) can be estimated in any of a number of ways. The diversity parameter used in the selection criterion is “group coancestry” ( $\Theta$ ), which is the probability that two genes taken at random from the gene pool, with replacement, are identical by descent (Cockerham 1967).

The selection criterion,  $B_\omega$ , for a set of genotypes,  $\omega$ , can then be formulated as

$$B_\omega = \bar{g}_\omega - c\Theta_\omega$$

where,  $c$  is a weighting constant,  $\bar{g}_\omega$  is the average of their breeding values, and  $\Theta_\omega$  is the average coancestry of the considered genotypes, which is a measure of their relatedness. Group-merit selection is simply a procedure to search for the set of  $\omega$  genotypes that maximises  $B_\omega$ .

A candidate for selection is penalised by the group coancestry that would result if it were included in the population together with the already selected members. The advantage a candidate offers is compared with other candidates for a particular stage of selection. Similarly, one can also compare individuals considered for substitution of already selected genotypes, and thus search for a set of members that maximises the group merit. The penalty weight ( $c$ ) can be set to give the desired balance between gain and relatedness. A penalty of zero would obviously maximise expected gain, while an infinite penalty would minimise relatedness and an infinitely negative penalty would maximise relatedness.

The search procedure applied by POPSIM to select a set of  $N$  individuals is:

1. Breeding values are predicted for each candidate, by conventional methods such as combined-index, BLUP, etc.
2. The penalty weight  $c$  is chosen by the user and supplied to POPSIM.
3. The subtraction term  $c\Theta_\omega$  for each family is calculated, in turn, by considering the resulting group coancestry of the selected population  $\omega$ , after the addition of another individual from this family. This term is subtracted from the predicted breeding values for each member of the family to estimate the group merit after the addition of that individual.
4. The candidate individuals are ranked by the selection criterion and the highest-ranking candidate is added to the selected population.
5. Steps 3 and 4 are repeated until  $N$  selections have been made, and the group merit of this set is recorded.

6. Steps 3 and 4 are repeated to add an  $N+1$ th individual to the selected set.
7. Each of the  $N+1$  selected genotypes is then removed, in turn, from the selected population and the group merit of the remaining  $N$  individuals calculated. The new set of  $N$  individuals giving the maximum group merit is identified.
8. If the group merit given by this new set of  $N$  individuals is less than before, the previous set of  $N$  individuals is retained and the selection algorithm is complete. Otherwise, this new set of  $N$  individuals is retained, and Steps 6 through 8 are repeated.

What this algorithm actually does is to first select  $N$  individuals consecutively, and then to substitute earlier selections by alternatively adding and discarding individuals until the process fails to find a set of  $N$  individuals with higher group merit. Note that a family will always be represented by its top-ranking phenotypes.