

Guideline to define breeding objective functions, construct selection indexes and deal with uncertain sires in sheep and goat breeding programmes



Joaquín Mueller¹, Tesfaye Getachew², Barbara Rischkowsky² and Aynalem Haile²

¹*National Institute for Agricultural Technology (INTA)*

²*International Center for Agricultural Research in the Dry Areas (ICARDA)*

June 2021

ISBN: 978-9291275212

Keywords: community-based breeding; sheep, goat, economic value of traits; Ethiopia

Manuals & Guidelines

ICARDA's Manuals & Guidelines series taps the Center's expertise to provide comprehensive advice and strategies that researchers can adopt to enhance agricultural productivity and overcome critical challenges affecting rural communities in the non-tropical dry areas.

About ICARDA

Established in 1977, the International Center for Agricultural Research in the Dry Areas (ICARDA) is a non-profit, CGIAR Research Center that focusses on delivering innovative solutions for sustainable agricultural development in the non-tropical dry areas of the developing world. We provide innovative, science-based solutions to improve the livelihoods and resilience of resource-poor smallholder farmers. We do this through strategic partnerships, linking research to development, and capacity development, and by taking into account gender equality and the role of youth in transforming the non-tropical dry areas.

Address

Dalia Building, Second Floor, Bashir El Kasser St, Verdun, Beirut, Lebanon 1108-2010. www.icarda.org

Disclaimer



This document is licensed for use under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/>

Unless otherwise noted, you are free to copy, duplicate, or reproduce and distribute, display, or transmit any part of this publication or portions thereof without permission, and to make translations, adaptations, or other derivative works under the following conditions:

-  **ATTRIBUTION.** The work must be attributed, but not in any way that suggests endorsement by the publisher or the author(s).
-  **SHARE ALIKE.** If this work is altered, transformed, or built upon, the resulting work must be distributed only under the same or similar license to this one

Cover photo — Candidate bucks selected based on their estimated breeding values and farmer's criteria, Konso, Ethiopia. (photo credit: ICARDA/Tesfaye Getachew).

Citation: Mueller J.P., Rischkowsky B., Getachew T., Haile A. 2021. Guideline to define breeding objective functions, construct selection indexes and deal with uncertain sires in sheep and goat breeding programmes, Addis Ababa, Ethiopia: ICARDA.

Copyright © 2021 ICARDA (International Center for Agricultural Research in the Dry Areas) All rights reserved.

Contents

List of Tables	i
List of Figures	ii
Acknowledgements.....	iv
Executive summary	v
Foreword.....	1
1. Definition of breeding objectives.....	2
1.1 Introduction	2
1.2 Breeding objectives: theoretical background	2
1.3 Goal traits.....	3
1.4 Economic importance of goal traits	3
1.5 Examples of breeding objectives.....	7
2. Construction and evaluation of selection indexes.....	11
2.1 Introduction	11
2.2 Selection indexes: theoretical background	11
2.3 Examples for the construction and evaluation of selection indexes	16
2.4 Computer tool to construct and evaluate unrestricted and restricted selection indexes	21
3. Estimation of breeding values in populations with uncertain sires.....	25
3.1 Introduction	25
3.2 Uncertain sire pedigree: theoretical background.....	25
3.3 Overview of the procedure	26
3.4 Work flow	26
3.5 Example of estimation of breeding values with uncertain sires.....	27
Supplementary material:	30
References	31

List of Tables

Table 1: Terminology used for discounted genetic expressions (DGE) of sires and dams in terminal and self-replacing herds for traits expressed in different moments of an animal’s life (Amer, 1999).....	7
Table 2: Phenotypic and genetic parameters in Merino sheep from Ponzoni (1979). Phenotypic correlations above diagonal, genetic correlations below diagonal.....	17
Table 3: Response to selection in economically important traits from single trait selection.	17
Table 4: Economic value, phenotypic and genetic parameters in sheep from Cunningham et al. (1970). Phenotypic correlations above diagonal, genetic correlations below diagonal.	20
Table 5: Index weightings for the regular index and the restricted index..	21

List of Figures

Figure 1: Constant herd dynamics assuming a known number of breeding dams (nD) and known biological parameters (from Mueller et al., 2021). 4	
Figure 2: Excel parameter file for the example of Ponzoni (1979).....	22
Figure 3: Selind output extract with input parameters of Figure 2.....	23
Figure 4: Capture of initial selindex.f95 statements showing run options for traits to be deleted, restricted or selected in stages.	24
Figure 5: Work flow to obtain BLUP breeding values from field data of populations with uncertain sires. In blue font are input/output files and in red font computer programs.....	27
Figure 6: Initial rows of example Excel file with typical field data and with up to three possible sires.	27

Acknowledgements

This guideline is an output of CRP Livestock. We greatly appreciate Sekota and Bonga Agricultural Research Centers and their staff for leading and implementing Abergelle goat and Bonga sheep community-based breeding programs. Special thanks to Bekahegn Wondimu, Yeshiwas Walle and Molatu Gobeze from Sekota and Zelalem Abate from Bonga Research Center for data management and clearing to be used as example in this guideline. The target-community members are acknowledged for cooperation during data collection and their participation in the whole phase of design and implementation of the breeding programs.

Executive summary

- 1) The first and crucial step in any breeding programme is to define the breeding goal. Small ruminant breeding programmes usually aim at simultaneous improvement of several traits. Guidelines are given to define multi-trait breeding objective functions by identifying goal traits and calculating their economic values using profit functions. Partial budgeting procedures are explained and exemplified with spreadsheet templates. The concept of discounted genetic expressions (DGE), which allow for goal trait expressions in future generations of a selection candidate, is introduced and a computer tool is provided for their calculation.
- 2) Genetic and economic improvement through selection of sires and dams for multi-trait breeding objectives require the choice of appropriate selection criteria, these are best formalized in selection indexes. Guidelines are given to construct selection indexes which improve or restrict genetic and economic gain in goal traits. Examples of index construction and index evaluation with different measurements on selection candidates are derived and replicated in spreadsheet files. A computer tool is provided to calculate the index weights and their accuracies, the relevance of traits included in the index, the expected genetic and economic gains in breeding goals as well as the outcomes when selection is in stages or when traits are dropped from the index.
- 3) In many low input breeding systems, the sire of a selection candidate is unknown and the true sire is known to be amongst a limited number of possible sires. In such cases, an uncertain sire relationship matrix can be used in the solution of the mixed models to predict breeding values and thereby increasing the accuracies of breeding values and increasing the eventual genetic gains. A computer tool is provided to construct the uncertain sire relationship matrix and a computer aided procedure with examples is described to calculate breeding values from field data in populations with up to three uncertain sires.

Foreword

In three main chapters this guideline describes basic procedures to (1) define breeding objectives, (2) construct and evaluate selection indexes and (3) estimate breeding values in populations with uncertain sires. The guideline targets practical sheep and goat advisors involved in the operation of animal breeding programmes, typically community-based programmes in low input systems. In such systems multiple goal traits are aimed, selection indexes are useful and incomplete pedigrees are common. After presenting the basic theoretical background to deal with each of these issues, the procedures are explained using field examples. Spreadsheet files and computer programs are provided as supplementary material to trace the steps involved in the calculations and to facilitate the application of the procedures to user defined cases.

1. Definition of breeding objectives

1.1 Introduction

The definition of the breeding objective is the first and crucial step in planning a selection program. Crucial, because efficient selection for the wrong objective may be worse than not selecting at all. Breeders are often interested in improving many traits. These traits may be of economical, biological, aesthetical or of any other nature. Breeders usually can easily identify the traits of interest but may have difficulties in determining the relative importance of each trait for selection emphasis. The economic importance of traits can be calculated using partial budgeting procedures. Some biological or aesthetical traits of interest can also be assigned an economic value but some may not. In what follows we assume that all traits of interest can be assigned an economic importance.

1.2 Breeding objectives: theoretical background

Let us consider the breeder's goal defined arbitrarily as H . It may be that H is the breeding value for a single trait, either one of X_1, X_2, \dots, X_n measured traits on the selection candidates or some other trait not measured. For m breeding objective traits in H we can write the following aggregate breeding objective function

$$H = f(Y_1, Y_2, \dots, Y_m),$$

where the Y_j 's are breeding values for economically important traits. Since selection produces only slow genetic changes we may use a linear approximation to find changes in H due to changes in Y , because the mean breeding value for any objective trait will not change greatly in one generation. Then, from a Taylor's series expansion of H about its value at current means we may write

$$H = \bar{H} + \frac{\partial H}{\partial Y_1}(Y_1 - \bar{Y}_1) + \frac{\partial H}{\partial Y_2}(Y_2 - \bar{Y}_2) + \dots + \frac{\partial H}{\partial Y_m}(Y_m - \bar{Y}_m),$$

where $\frac{\partial H}{\partial Y_j}$ is the partial derivative of H with respect to Y_j evaluated at $(\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_m)$.

Less formally, $\frac{\partial H}{\partial Y_j}$ is the rate of change in the average breeding objective for a small change in Y_j when all other traits remain unchanged. The rate of change can be estimated numerically by calculating \bar{H} with current \bar{Y}_j values, then recalculating \bar{H} with a change $\Delta\bar{Y}_j$ in only one trait, to find the marginal change $\Delta\bar{H}$. Then approximately $\frac{\partial H}{\partial Y_j} = \frac{\Delta\bar{H}}{\Delta\bar{Y}_j}$.

These marginal changes in \bar{H} per unit changes in goal traits of economic importance are usually referred to as *economic values* and may be denoted $EV_{Y_1}, EV_{Y_2}, \dots, EV_{Y_m}$ (often also denoted as a_1, a_2, \dots, a_m). Apart from an additive constant we may then write for aggregate breeding value

$$H = EV_{Y_1}Y_1 + EV_{Y_2}Y_2 + \dots + EV_{Y_m}Y_m.$$

Sometimes we may want to judge the similarity of breeding objective functions with the same traits but different economic values. Suppose we have two different objectives, H_1 and H_2 , with the same traits and but vectors of different economic values \mathbf{a}_1 and \mathbf{a}_2 , then their genetic correlation is $r_{H_1H_2} = \mathbf{a}_1' \mathbf{Q} \mathbf{a}_2 / \sqrt{(\mathbf{a}_1' \mathbf{Q} \mathbf{a}_1)(\mathbf{a}_2' \mathbf{Q} \mathbf{a}_2)}$, where \mathbf{Q} is the covariance of traits in H (James, 1982). Note that symbols in bold lower case represent vectors and symbols in bold upper case font represent matrices here and throughout this guideline.

1.3 Goal traits

In order to determine traits of economic importance we first look at sources affecting returns and costs in the herd. These sources affecting profit will depend on the particular breed and production system. Returns may be due to meat production, milk production or wool production and costs may be due to feed, labour or marketing expenses. Returns from meat production can be specified as due to sale of surplus lambs and sale of cast-for-age adults. We then look at animal traits influencing these sources of profit. For example, returns from the sale of lamb meat are influenced by traits such as the number of lambs born per ewe, survival of lambs from birth to sale age, sale weight and carcass quality traits. Similarly, returns from wool production are influenced by fleece weight and fleece quality traits and returns from milk production are influenced by average daily milk yield, lactation length, and milk quality traits. Costs due to feed expenses may be related to a trait such as food intake and costs due to labour or veterinary expenses may be related to easy care traits such as face wool-cover grade or faecal-egg count.

All traits influencing profit should be considered as goal traits even if they cannot be assessed on selection candidates. The breeding objective must be defined in terms of what the breeder would like to improve and not in terms of what he can measure. This is because if goal traits have genetic correlations with traits in the selection criterion (the selection index), their means will usually be altered by selection and these alterations need to be controlled.

1.4 Economic importance of goal traits

Once we have identified the goal traits their economic values can be derived from profit functions. Profit due to a goal trait being the difference between revenues R and variable costs C . With Y_1, Y_2, \dots, Y_m traits contributing to overall profit we can write overall profit due to animal products as

$$P = \sum_{i=1}^m f_i \times (R_i - C_i) \bar{Y}_i - K.$$

In this equation, profit due to Y_i is weighted by the frequency f_i or number of expressions of the trait in a chosen time unit. For example, in an ewe's lifetime there will be several lactation expressions but only one cull-for-age ewe sale-weight expression. The economic value of a trait is then calculated as the partial derivative of P with respect to that trait evaluated at the average value for all other traits as explained previously (section 1.2 Theoretical background). For example, the economic value of Y_i will be $EV_{Y_i} = \partial P / \partial Y_i$. The fixed overhead cost K

vanishes on differentiation. Note that the result of the differentiation is equivalent to $P^* - P$, where P^* is P with a unit increase in trait Y_i , or simply $R_i - C_i$ per unit of the trait.

We have to look more closely on the time unit chosen to calculate trait expressions. To make profits comparable, the same time unit must be used for all traits. A convenient and simple option is to calculate expressions of traits per year. Expressions of traits per year can be calculated from average field data or from a suitable model representing the herd and production system of interest. Figure 1 illustrates a sheep or goat model which allows calculating expressions of various traits in a herd with a given number of dams (nD) and known biological parameters.

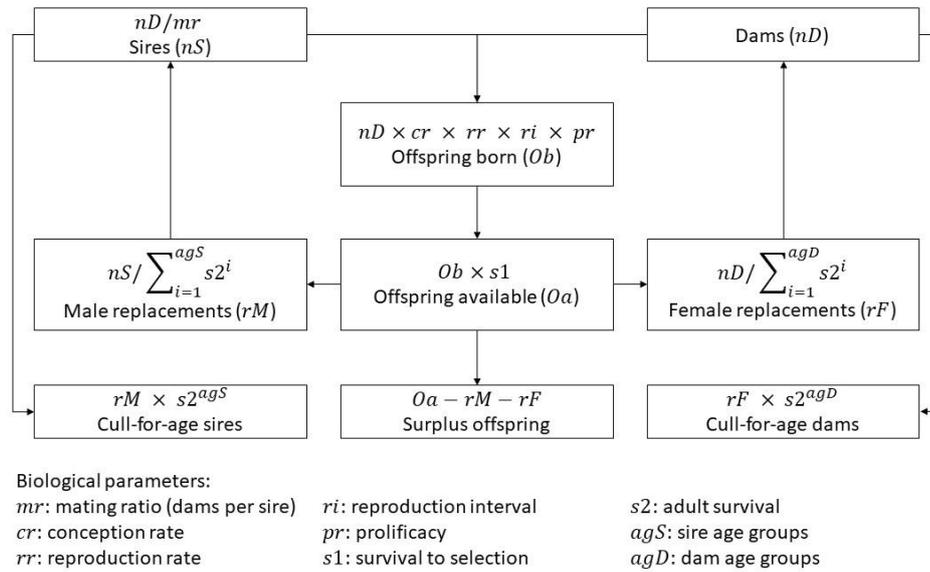


Figure 1: Constant herd dynamics assuming a known number of breeding dams (nD) and known biological parameters (from Mueller et al., 2021).

The previous method of calculating economic values accounts for frequencies of trait expression per year but does not consider the timing of trait expression. Since revenues and costs of different traits usually accrue at different times, discounting is appropriate to account for different opportunity costs. For example, cashing on lamb sales occurs before cashing on cast-for-age sales. Therefore, future expressions of cast-for-age weight should in principle be more discounted than future expressions of weaning weight. This can be done by including discount rates D and the previous profit equation. Discounting factors would normally be the same for revenues and costs, but some authors prefer specific discounting factors D and D^1 so that following Ponzoni (1982) we have

$$P = \sum_{i=1}^m f_i \times (R_i D_i - C_i D_i^1) \bar{Y}_i - K.$$

While this equation considers frequency and timing of the expression of a trait it does not consider expressions in future generations. When selecting a ram on breeding values for

different traits the economic values should consider the number of expressions of each trait in his progeny and in their descendants. McClintock and Cunningham (1974) developed a procedure using gene flow techniques and discounting to calculate what they called “discounted genetic expressions” (DGE). The calculation of DGEs is somewhat awkward but Amer (1999) presented a method with practical formulae in matrix notation to calculate DGEs of different traits for rams and ewes in terminal or self-replacing herd situations as will be explained later in the Example 3. Discounted genetic expressions of trait Y_i can be multiplied by $(R_i - C_i)$ to obtain its discounted economic value in given time horizons and discount rates.

Feed costs

Feed intake or residual feed intake should be a goal trait in most selection program situations in order to account for feed costs. The scarcity of genetic and phenotypic parameters involving these traits as well as the difficulties to measure these traits, particularly in grazing animals lead most authors to consider feed costs indirectly. Typically, additional feed intake costs due to increased meat or milk production are subtracted from the revenues due to meat and milk production traits. The weakness of this approach, is that it assumes that feed consumption is exclusively a function of increased growth, reproduction or milk production rates, and that it has no genetic variance independent of those traits.

When using the reduced revenue approach, marginal increases of feed costs can be considered proportional to body weight. For example, if the annual feed cost for a 50 kg ewe is \$ 1.50 then the annual feed cost for a 51 kg ewe is $1.50 + 1.50/50 = \$ 1.53$. Ponzoni (1979) assumed feed maintenance requirements to be proportional to metabolic weight (body weight to the power of 0.75) rather than to body weight. In this case, the annual feed cost of a 51 kg ewe results in $1.50 + 1.50 \times (51^{0.75} - 50^{0.75})/50^{0.75} = \$ 1.52$. Kosgey et al. (2003) and other authors calculated feed costs by considering metabolizable energy (ME) requirements for maintenance and production (growth, reproduction and lactation) and then, with various assumptions, they related ME to dry matter (DM) and DM to fresh roughage with its market price.

The previous feed cost calculations were based on a constant feed supply basis. An alternative way of dealing with feed costs is to translate increased feed costs into a reduction of feed availability by reducing the carrying capacity of the pasture. In that case instead of increasing feed costs, fewer animals entail fewer trait expressions. It should be noted however that increased feed intake may be irrelevant for a producer who is understocked or whose herd lambing period coincides with a period of surplus pasture production. This might also be the case in pastoral conditions where little amount of feed is purchased and land is not the main restriction of flock size as often is the case with nomadic farmers. On the other hand, profit in intensive farming systems based on the regular purchase of feedstuff may depend heavily on changes in feed costs. Researchers need to evaluate the option to deal with feed costs considering the particular production system and information available.

Discounting factors

Discount rates affect overall profit and change relative economic values of traits expressed in different time. High discounting factors shifts economic values to early expressed traits. Farm mortgage interest rates can be used as a guide for the choice of discount rates but the implicit inflation rate should be subtracted. Inflation rate affects both, prices and costs and therefore should not be included in the discount rate. Discounting factors will also depend on the type of breeding program in consideration. In general, company programs would consider higher discount factors than governmental or community programs. In any case regular discounting rates usually do not substantially change relative trait importance in programs with goal traits expressed periodically like lambing, milking and shearing in sheep.

Combining income and expenses in different ways

In this guideline economic values were calculated as profit (P) = income – expenses (or revenues – costs). Other ways of combining income and expenses are return to investment (R) = income / expenses and cost per unit production (Q) = expenses / income. Ponzoni (1988) compared breeding objectives using P , R and Q equations concluding: (i) that combining income and expense as income / expense (R) gave the same result as combining them as expense / income (Q); (ii) that if profit is equal to zero (a case when an industry wide equilibrium economic profit is included as a cost in the accountant sense) the same economic values will be obtained whether income and expense are combined as a difference or as a ratio, and (iii) even when profit is not equal to zero the effect on selection decisions of combining income and expense in different ways is likely to be negligible. It is recommended to refer to Ponzoni (1988) for a more detailed discussion on the implications using different ways of calculating economic values. In practice, combining income and expense as a difference or as a ratio would have only a small effect in a breeding program, because usually profit is small relative to total costs, and because selection indexes are robust to moderate changes in economic values.

Dealing with traits without economic information

A bulk of research has been done on breeder preference trait identification and its relative importance, effectively allowing to apply the previous procedures to define a breeding objective function. More difficult is the economic valuation of minor malformations of jaws, backs, legs and feet. It is difficult to obtain accurate information about them because often they cannot be assessed objectively, and also because some culling against animals showing the defect is constantly taking place in most herds anyway. In practice, provided the amount of culling on the basis of these traits is small, ignoring these traits in the objective function is unlikely to affect seriously the effectiveness of a breeding programme. As for the elimination of a few aesthetically unsatisfactory individuals, it may rather contribute positively towards the breeder's interest in the breeding programme. Thus, a two-stage selection procedure involving selection on breeding values followed by farmers' selection among top-ranking candidate sires is usually recommended.

1.5 Examples of breeding objectives

Example 1: Breeding objective for wool sheep

This example is based on a profit equation for Merino sheep published by Ponzoni (1986). Economic values were derived for situations with and without discounting. Here we present the methods applied for the case without discounting. The example is replicated in an Excel file¹ which allows to follow the derivation of number of trait expressions, incomes and costs. The production system studied is a breeding flock in which incomes are generated by sale of wool, sale of all surplus offspring as hoggets (aged about 14 months), and sale of cull-for-age ewes. Rams are supposed to be purchased and are not included in the analyses. Expenses are related to feeding, husbandry, wool harvesting and marketing of wool and surplus sheep.

Example 2: Breeding objective for meat sheep and dairy goats

This example is based on profit equations for Abergelle goats and Bonga sheep in community-based breeding programs published by Mueller et al. (2021). The procedures are similar to the previous example with corresponding traits of interest, biological parameters and economical parameters. The example is replicated in an Excel file².

Example 3: Breeding objective using discounted genetic expressions (DGE)

Examples 1 and 2 were based on herd profit per year and did not consider trait expressions in further descendants of selection candidates nor time lag of trait expression. In this example the technique of discounted genetic expressions (DGE), which accounts for both, expression in future descendants and time lags is explained using Amer's (1999) approach and example. The procedure is used to calculate various DGE of interest in sheep and goat programs. For example, in community-based breeding programs when selecting sires for self-replacing local breeds the DGE for a trait like lamb survival is identified as MRL, slaughter weight as MRS, hogget fleece weight as MRH, number of lambs born as MRA and cull-for-age weight as MRC (Table 1).

Table 1: Terminology used for discounted genetic expressions (DGE) of sires and dams in terminal and self-replacing herds for traits expressed in different moments of an animal's life (Amer, 1999).

Trait expression	DGE for sires			DGE for breeding females	
	Terminal Sire	Sire to breed terminal ewes	Breed self-replacing ewes	Terminal	Self-replacing
Lamb at birth	MTL	MEL	MRL	FEL	FRL
Lamb at slaughter	MTS	MES	MRS	FES	FRS
Hogget		MEH	MRH	FEH	FRH
Ewe (annual)		MEA	MRA	FEA	FRA

¹ Excel file [Ponzoni 1986 breeding objective.xlsx](#) available as supplementary material.

² Excel file [Mueller et al 2021 breeding objective.xlsx](#) available as supplementary material.

Ewe (at culling)	MEC	MRC	FEC	FRC
------------------	-----	-----	-----	-----

To calculate DGE for sheep (or goats) the following parameters are needed:

n: highest ewe age (in years) considered possible

s: vector of n probabilities that an ewe survives and lambs from age $i-1$ to age i (between 0 and 1.0). When ewes lamb first at 2 years of age $s_1=0$

p: vector of n numbers of lambs reaching slaughter or reproductive age per ewe lambing at age i (between 0 and 1.0). When ewes lamb first at 2 years of age $p_1=0$

c: cull-for-age threshold (must be equal or smaller than n)

h: years in planning horizon (between 1 and 50)

r: discount rate (between 0 and 1.0)

y: number of years a sire is used for mating

e: number of ewes mated per sire

u: proportion of ewe lambs retained (between 0 and 1.0)

ls: probability of survival from birth to slaughter or reproduction (between 0 and 1.0).

The matrix operations involved to obtain the various DGE with these parameters were developed by Amer (1999). Here we use a convenient computer program³ called **AmerDGE.for** developed by Prof. John James (University of Sydney) which calculates the DGE of Table 1. To run the program an input parameter file must be provided. For each set of DGE calculations, parameters shall be submitted in four lines and separated by spaces in the following sequence:

```
n
s1 s2 ... sn
p1 p2 ... pn
c h r y e u ls
```

To run the program, run the command prompt and change to the directory which contains the program. Then type AmerDGE (any combination of upper and lower case letters will do). The program asks "How many sets of DGE are required?" and there needs to be a set of parameters on four lines in the input file for each set of DGE required. The program then asks "What is the input file name?" If the input file is not in the same directory as the program, the path must be given. The program next asks "What is the output file name?" and expects this name to be for a new file. Make sure that the directory does not contain a file of the name supplied.

To illustrate the use of the computer program the example with base parameters from Amer (1999, Table 4) is replicated. The parameter file⁴ is then as follows:

³ Computer program **AmerDGE.for** supplied as supplementary material.

⁴ Example parameter file **inAmer** supplied as supplementary material.

```

7
0 0.95 0.94 0.93 0.88 0.83 0.70
0 1.10 1.25 1.25 1.35 1.30 1.10
5 10 0.07 3 90 0.85 0.8

```

The output file⁵ will contain the published results of Amer (1999, Table 5 for $r=0.07$) as follows:

FEL	FES	FEH	FEA	FEC
2.22	1.78	1.00	2.90	0.78
MTL	MTS			
181.51	145.21			
MEL	MES	MEH	MEA	MEC
309.54	185.92	57.68	167.45	45.05
FRL	FRS	FRH	FRA	FRC
3.21	1.95	1.54	4.21	1.06
MRL	MRS	MRH	MRA	MRC
366.79	196.02	88.66	243.00	61.40

The same results are also obtained in an Excel file⁶ with the dimension $n = 7$ of the example. For other dimensions the Excel file must be rearranged completely or rather the computer program should be used.

Appropriate DGE of m traits each multiplied by its economic value (EV) and the animal's breeding value (BV) gives aggregate breeding value H as:

$$H = \sum_{i=1}^m DGE_i \times EV_i \times BV_i.$$

Amer (1999) presents the following case for use of the resulting DGE. Consider a purchase decision by a commercial farmer choosing between two rams, π^1 and π^2 , with some daughters of the rams expected to produce future flock replacements. The two rams are similar in all traits except in number of lambs born per ewe lambing (NLB) and lamb survival (LS), for which estimated breeding values (BV) are available. To compare the rams, economic values for a single expression (one lambing year) of number of lambs born (EV_{NLB}) and for a single expression (per lamb born) of lamb survival (EV_{LS}) are required. The difference between the two rams in net present value to the farmer ($\pi^1 - \pi^2$), is calculated as

$$\pi^1 - \pi^2 = (BV_{NLB}^1 - BV_{NLB}^2) \times EV_{NLB} \times MRA + (BV_{LS}^1 - BV_{LS}^2) \times EV_{LS} \times MRL,$$

⁵ Example output file [outAmer](#) supplied as supplementary material.

⁶ Excel file [Amer 1999 DGE.xlsx](#) supplied as supplementary material.

where the BV and π superscripts correspond to sire (1 and 2), MRA is the number of DGE for a ewe lambing trait (expressed annually), and MRL is the number of DGE for a lamb trait (per lamb born) expressed pre-weaning.

2. Construction and evaluation of selection indexes

2.1 Introduction

The use of some combination of traits in a linear “index” to rank animals for selection has been and still is popular. For example, an index I given as ten times fleece weight (FW) plus body weight (BW) to select wool sheep $I = 10 \times FW + BW$. The logic behind such an index being that young ram candidates often have approximate fleece weights of 4 kg and body weights of 40 kg and if equal importance is set to each trait the two measurements could be scaled and added for selection purpose. However, such an index involves several assumptions and is not necessarily efficient. There are many other indexes combining traits in different formulas. For example, the “base” index, whereby phenotypic values of, say n measured traits (X), are multiplied by their relative importance or economic value (EV), so that $I = \sum_{i=1}^n EV_i \times X_i$. Since such an index ignores differences in heritability (h^2) of traits, an “heritability index” has been proposed $I = \sum_{i=1}^n EV_i \times h_i^2 \times X_i$. However, this index ignores correlations between traits and assumes all traits of interest are measured. We will refer here to the so called “optimum”, “standard” or “Smith - Hazel” index named after the authors who introduced its derivation. This standard index maximises response to selection given a set of measurements and appropriate parameters.

As will be seen later if BV for all goal traits are available and if these BV were obtained from a multivariate BLUP analyses including all goal traits, selection can be performed directly on aggregate breeding value $H = \sum_{i=1}^n EV_i \times BV_i$ where BV_i is the BLUP multivariate breeding value for trait i . When multivariate BLUP breeding values are not available standard index selection can be very useful (Mueller et al., 2021).

2.2 Selection indexes: theoretical background⁷

Construction of the standard selection index to improve H

Suppose we have measurements on n traits for each animal in a population. Denote the traits by X_1, X_2, \dots, X_n . Suppose we wish to improve breeding value, denoted by H , as quickly as possible. We wish to do this by using an index I combining the records on the n traits. We write $I = b_1X_1 + b_2X_2 + \dots + b_nX_n$. H may be defined arbitrarily; it is the breeder’s goal. If H and X_i have a multivariate normal distribution the following theory is exact. Otherwise it is approximate, since all regressions are assumed linear. If I and H have standard deviations σ_I and σ_H and i is the standardised selection differential, the gain per generation in H following selection on I is $R_{H,I} = i\sigma_I b_{HI} = ir_{IH}\sigma_H$. Here b_{HI} is the regression of H on I and r_{IH} is the correlation between I and H . Since i and σ_H do not depend on which index is used for selection, to get maximum response we must maximise r_{IH} . There is no unique solution to this problem since if we multiply all index values by the same constant the correlation is unaltered. Thus we may choose the scale of the index at will.

⁷ This section is largely based on lectures of Prof. John James (University of Sydney).

Let P_{ij} denote the covariance of X_i and X_j (P_{ii} is the variance of X_i) and g_{jH} denote the covariance of X_j and H . Then it can be shown that the values of b_1, b_2, \dots, b_n which maximise r_{IH} are given by the solutions of the following equations

$$P_{11}b_1 + P_{12}b_2 + \dots + P_{1n}b_n = g_{1H}$$

$$P_{21}b_1 + P_{22}b_2 + \dots + P_{2n}b_n = g_{2H}$$

...

$$P_{n1}b_1 + P_{n2}b_2 + \dots + P_{nn}b_n = g_{nH}.$$

In matrix notation this is $\mathbf{Pb} = \mathbf{g}$ and the solution can be written $\mathbf{b} = \mathbf{P}^{-1}\mathbf{g}$. So the procedure is to solve the equations to find b_1, b_2, \dots, b_n and use these values together with performance records to calculate an index value for every animal, selecting those with the best index values.

Variance and accuracy of index

The variance of the index is given by

$$\sigma_I^2 = b_1^2 P_{11} + b_2^2 P_{22} + \dots + b_n^2 P_{nn} + 2b_1 b_2 P_{12} + 2b_1 b_3 P_{13} + \dots + 2b_{n-1} b_n P_{n-1,n}$$

or in matrix notation $\sigma_I^2 = \mathbf{b}'\mathbf{Pb} = \mathbf{g}'\mathbf{P}^{-1}\mathbf{g}$.

The covariance of H and I is

$$\text{cov}(HI) = b_1 g_{1H} + b_{21} g_{2H} + \dots + b_n g_{nH} = \mathbf{b}'\mathbf{g} = \mathbf{g}'\mathbf{P}^{-1}\mathbf{g} = \sigma_I^2.$$

Thus, with this choice of scale for the index $b_{HI} = \text{cov}(HI)/\sigma_I^2 = 1$. That is, the regression of breeding value on index is 1. The gain per generation is $i\sigma_I$. Thus the standard deviation of the index measures the rate of response to selection. The correlation is $r_{HI} = \sigma_I/\sigma_H$.

Now suppose H is a function of m goal traits Y_1, Y_2, \dots, Y_m and denote the genetic covariance between goal traits Y_i and index traits X_j by G_{ij} . Then it follows that extending g_{jH} we have $g_{jH} = EV_1 G_{1j} + EV_2 G_{2j} + \dots + EV_m G_{mj}$. Thus, the index coefficients are solutions of the equations

$$P_{11}b_1 + P_{12}b_2 + \dots + P_{1n}b_n = EV_1 G_{11} + EV_2 G_{12} + \dots + EV_m G_{1m}$$

$$P_{21}b_1 + P_{22}b_2 + \dots + P_{2n}b_n = EV_1 G_{21} + EV_2 G_{22} + \dots + EV_m G_{2m}$$

...

$$P_{n1}b_1 + P_{n2}b_2 + \dots + P_{nn}b_n = EV_1 G_{n1} + EV_2 G_{n2} + \dots + EV_m G_{nm}.$$

In matrix terms this is $\mathbf{Pb} = \mathbf{Ga}$ where \mathbf{P} is an $n \times n$ matrix, \mathbf{b} is an n vector, \mathbf{G} is an $n \times m$ matrix and \mathbf{a} is an m vector of economic values. The solutions $\mathbf{b} = \mathbf{P}^{-1}\mathbf{Ga}$ are the previous results with $\mathbf{g} = \mathbf{Ga}$.

Combining sub-indexes

The problem can be approached a little differently. We could develop a series of indexes, one for each of the economically important traits. These would give the best estimates of breeding value for each trait. Thus we could have an index I_1 to maximise genetic gain in X_1 , an index I_2 to maximise gain in X_2 , an index I_m to maximise gain in X_m . We can then combine these separate indexes into one combined index $I = EV_1I_1 + EV_2I_2 + \dots + EV_mI_m$. This index will be the same as that derived above. The separate indexes can be obtained using the general procedure setting all except one of the EV_j equal to zero. In matrix notation, the coefficients for the individual trait indexes are given by the columns of the matrix $P^{-1}G$. It is sometimes considered useful to calculate these coefficients in order to be able to modify the combined index simply if different economic values are required.

Indexes with different amount of information

In practice there will be selection candidates with partial or missing data. If there are many such animals, appropriate indexes should be constructed with the information available. If there are only a few animals with missing data, the contemporary group mean can be assigned as their expected value. More conveniently, all measurements can be recorded as deviations from the mean in which case missing values will be zero. It may also be convenient to adjust the measurements to means and variances actually used in the derivation of the particular index. For example, field measurement X with mean \bar{X} and standard deviation σ_X can be standardized to the population mean μ and population standard deviation σ used to obtain index weights as $(X - \bar{X})/\sigma_X \times \sigma + \mu$. However, if all animals have equal information, actual values may be used in the calculation of an animal's index.

Information of relatives in an index

Information on relatives may be included in an index. For example, the information available on an animal could be: its own weaning weight, its own yearling weight, its own fleece weight at first shearing, number of lambs weaned (NLW) by its dam at 3 lambings and dam's fleece weight at first shearing. An index can be derived using the general method, with appropriate variances and covariances for these traits making allowance for the genetic relationship between the animal being considered for selection and the relative providing the performance records. For example, in the case of dams NLW, only $\frac{1}{2}$ of the variances and covariances with NLW should be used. Likewise, if half-sib information is used in the index, only $\frac{1}{4}$ of the variances and covariances should be used. It may be useful, in these cases, to redefine the trait measured. For example, dams number of lambs weaned as dNLW with the parameters for NLW already divided by 2.

Restricted selection index

Selection indexes can also be constructed to satisfy certain restrictions. For instance, if one or two traits are already at optimum values, e.g. ewe weight, it is possible to devise an index which will hold these traits constant while changing other traits as rapidly as possible in the desired directions. Kempthorne and Nordskog (1959) introduced the idea of restricting gain in a trait using the following reasoning. If breeding value of trait Y_i is restricted to zero

then its economic value is zero, $EV_{Y_i} = 0$ and the covariance between index and breeding value must be zero, $cov(I, Y_i) = 0$. A Lagrange multiplier or dummy trait equation is then added to \mathbf{P} . The solution is as usual $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$. Cunningham et al. (1970) proposed an equivalent method to do this by augmenting the basic unrestricted equations $\mathbf{P}\mathbf{b} = \mathbf{G}\mathbf{a}$ as follows: add a dummy variable to the index, a row and column to \mathbf{P} , and a row of zeros to \mathbf{G} . The resulting equations are

$$\begin{bmatrix} \mathbf{P} & \mathbf{G}_i \\ \mathbf{G}'_i & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{b}_{dummy} \end{bmatrix} = \begin{bmatrix} \mathbf{G} \\ \mathbf{0} \end{bmatrix} \mathbf{a}$$

or $\mathbf{P}^*\mathbf{b}^* = \mathbf{G}^*\mathbf{a}$.

The $(n + 1)$ th column of \mathbf{P}^* consists of the j th column of \mathbf{G} for its first n elements, with a zero in the final position. This last trait would be the trait restricted to remain with no genetic change. The restricted index weightings are then found as usual $\mathbf{b}^* = \mathbf{P}^{*-1}\mathbf{G}^*\mathbf{a}$. The method can be extended to restrict more traits by adding corresponding rows and columns. The method can also be adapted to produce desired levels of restriction by replacing the $\mathbf{0}$ in \mathbf{P}^* with a negative constant and testing desired restriction levels by trial and error. An alternative way of restricting change in a trait is to accommodate the economic value of the trait to be restricted by trial and error until the desired level of restriction is attained. An interesting result of Cunningham's method described above is that \mathbf{b}_{dummy} gives the negative economic value which produces zero gain when using the unrestricted selection index.

It should be noted that restricting a trait to zero change should always be proposed with care since undesired correlated changes may be forced and overall economic benefit may be seriously affected. James (1982) noted that however useful restricted indexes may be, they are in a sense a reversion to the concept of an "ideal animal" as used by traditional breeders. In effect, by restricting certain traits to "optimum" values, the breeder is giving infinite negative economic value to deviations from these optima, because it is implied that no extra gain in other traits which could be achieved by allowing changes in the restricted traits could possibly compensate for the loss caused by these changes. Since this is unlikely to be true, it would seem preferable to use the unrestricted index, recalculating economic values as means change or using quadratic approximations to total merit.

Desired gains index

Breeders may be in trouble assigning economic values to traits but may be willing to specify the amount of gain they would like to see in each trait of interest. Desired gains can be obtained using procedures developed by Pesek and Baker (1969) who modified the standard selection index by replacing economic values with desired gains. The procedure is basically an extension to restricted selection indexes. The objective is to find the linear index $I_d = \mathbf{b}'_d\mathbf{z}$ giving a pre-specified vector of proportional responses in each character. Besides providing control over how each individual character changes, the desired-gains index does not require specification of economic values \mathbf{a} . It does, however, still require estimates of \mathbf{P} and \mathbf{G} . Following Walsh and Lynch (2000) let Δ_d denote the vector of desired changes, so that the ratio of any two elements, Δ_i/Δ_j is the desired ratio of response in characters i and j .

Selection $I = \mathbf{b}'\mathbf{z}$ gives a vector of response proportional to $\mathbf{G}\mathbf{b}$. Hence $\mathbf{G}\mathbf{b}_d = \Delta_d$, giving the vector of weights for the desired-gains index as $\mathbf{b}_d = \mathbf{G}^{-1}\Delta_d$.

The equation assumes that desired response is specified for all n measured traits. More generally, suppose there are n traits in the phenotypic index and $m < n$ traits whose responses are of interest. Denote by \mathbf{G}_r the $m \times n$ matrix whose ij -th element denotes the additive genetic covariance between measured trait i and response trait j . Solutions are still of the form $\mathbf{G}_r\mathbf{b}_d = \Delta_d$, but this does not yield a unique solution for \mathbf{b}_d as \mathbf{G}_r is not a square matrix and hence \mathbf{G}_r^{-1} is not defined. A unique solution can be obtained by imposing the standard constraint of maximizing response for a fixed amount of selection. For a fixed selection intensity \bar{i} , response increases as $\sigma_I^2 = \mathbf{b}'\mathbf{P}\mathbf{b}$ decreases. A unique solution can thus be obtained by solving $\mathbf{G}_r\mathbf{b} = \Delta_d$ subject to the additional constraint that $\mathbf{b}'\mathbf{P}\mathbf{b}$ is minimized, giving $\mathbf{b}_d = \mathbf{P}^{-1}\mathbf{G}'_r(\mathbf{G}_r\mathbf{P}^{-1}\mathbf{G}'_r)^{-1}\Delta_d$. As expected, this reduces to $\mathbf{b}_d = \mathbf{G}^{-1}\Delta_d$ when $\mathbf{G}_r = \mathbf{G}$. A published example of the procedure can be found in Kargar Borzi et al. (2017). A worked example of the construction of a desired gain index is provided as an Excel file⁸.

Response to index selection

The responses to index selection can be worked out without much difficulty. The genetic gain in trait Y_j is $i\sigma_I b_{Y_j, I} = i \text{cov}(Y_j, I)/\sigma_I$. We have already seen how to calculate σ_I . Also $\text{cov}(Y_j, I) = b_1 G_{1j} + b_2 G_{2j} + \dots + b_n G_{nj}$. The changes in mean breeding value for all traits can then be calculated. The contribution of change in the j -th trait to economic gain is then $a_j [i \text{cov}(Y_j, I)/\sigma_I]$. Total economic gain $i\sigma_I$ is then the sum of these contributions.

Efficiency of selection criteria included in the index

If we wish to compare the efficiency of different selection criteria we can calculate r_{IH} or σ_I for each criterion and thus see the relative efficiencies of different indexes. A simpler procedure can be used to assess the effect of dropping a single trait from the index if we have computed the matrix $\mathbf{R} = \mathbf{P}^{-1}$. Let I^* be the index obtained when the k -th trait is omitted from the full index I . The efficiency of the reduced index relative to the full index is $\sqrt{1 - b_k^2/R_{kk}\sigma_I^2}$. This result makes it relatively easy to assess whether dropping any trait from the index will seriously affect response to selection. The value of the trait in the index or percent reduction in genetic gain for aggregate genotype if the trait is omitted is then $1 - \text{efficiency}$.

The variance of the reduced index is $\sigma_{I^*}^2 = \sigma_I^2 - b_k^2/R_{kk}$ and the coefficients in the reduced index are $b_j^* = b_j - b_k R_{jk}/R_{kk}$.

⁸ Excel file [Walsh and Lynch 2000 desired gain selection index.xlsx](#) available as supplementary material.

Indexes using BLUP breeding values

Sometimes BLUP breeding values are combined in an overall economic index. The principles for calculating the weights of these indexes are similar to those outlined in the previous sections. When BLUP breeding values are from a complete multivariate analysis the optimal index weights (\mathbf{b}) are the sum of the partial regressions coefficients of each goal trait on each index trait weighted by the economic value of the goal trait. Given m traits in the selection goal and n traits in the index the partial regressions can be calculated as $\mathbf{R} = \mathbf{G}^{-1}\mathbf{G}_{ig}$ and $\mathbf{b} = \mathbf{R}\mathbf{a}$ where \mathbf{R} is a matrix of partial regressions \mathbf{G}_{ig} is the matrix of genetic covariance between m goal traits and \mathbf{G} is the covariance matrix between the index traits and \mathbf{a} is the vector of economic values.

When goal and index traits are the same, $\mathbf{G} = \mathbf{G}_{ig}$ then $\mathbf{R} = \mathbf{1}$ and $\mathbf{b} = \mathbf{a}$ (Mrode, 2005). In other words, BLUP breeding values simply need to be weighted by their economic value. If goal and index traits are not the same, \mathbf{R} can be calculated from regression of phenotype on the estimated breeding values for the index traits. However, if BLUP breeding values are from univariate analysis rather than from multivariate analysis the use of \mathbf{a} as weighting factor depends on the size of the ignored correlations between traits.

2.3 Examples for the construction and evaluation of selection indexes

Example 2.1: Selection index for sheep

This example describes the calculations necessary to obtain selection index weights. Genetic gain from index selection is compared to genetic gain from single trait selection. The calculations are also replicated in an Excel file⁹.

Suppose the breeding objective function or aggregate breeding value H published by Ponzoni (1979). The traits of interest are clean fleece weight (CFW), mean fibre diameter (MFD), number of lambs weaned (NLW), weaning weight (WWT) and ewe body weight (EBW) with economic values (EV) of genetic gains per ewe lifetime in Australian Dollars (\$). Thus,

$$H = \$ 10.7 \times CFW - \$ 1.5 \times MFD + \$ 53.9 \times NLW + \$ 0.94 \times WWT + \$ 0.13 \times EBW.$$

The minus sign for MFD indicates negative economic gain when MFD increases.

Suppose that selection in a flock can be carried out to give standardised selection differentials of $i_M = 2.3$ in rams and $i_F = 0.7$ in ewes, so that the average standardised selection differential is $i = 1.5$. Suppose also an average generation length of $L = 3$ years.

Let us first estimate the responses in economically important characters and overall economic gain per year if single trait selection is practised for yearling or hogget body weight (YBW), greasy fleece weight (GFW) and low wrinkle score (WRS). The necessary genetic and phenotypic parameters used in the construction of the desired selection index are in Table 2.

⁹ Excel file [Ponzoni 1979 selection index.xlsx](#) available as supplementary material.

Note that none of the measured traits are actually goal traits, but note also that they are genetically correlated (Table 2).

Table 2: Phenotypic and genetic parameters in Merino sheep from Ponzoni (1979). Phenotypic correlations above diagonal, genetic correlations below diagonal.

	CFW kg	MFD mic	NLW	WWT kg	EBW kg	GFW kg	YBW kg	WRS score 0-13
Mean	3.5	23.0	0.78	30.0	50.0	5.5	48.0	4
h^2	0.4	0.5	0.1	0.25	0.4	0.35	0.4	0.5
σ_p^2	0.25	4.7	0.22	9.0	20.25	0.28	20.25	1.5
CFW		0.2	0.0	0.25	0.3	0.85	0.3	0.2
MFD	0.25		0.1	0.1	0.13	0.13	0.13	0.14
NLW	0.0	-0.1		0.12	0.15	0.0	0.15	-0.2
WWT	0.25	0.1	0.2		0.35	0.25	0.35	-0.08
EBW	0.2	0.1	0.25	0.7		0.3	0.7	-0.08
GFW	0.75	0.16	0.0	0.25	0.2		0.3	0.3
YBW	0.2	0.1	0.25	0.9	0.8	0.2		-0.03
WRS	-0.15	0.18	-0.3	-0.25	-0.25	0.2	-0.25	

Economic value of genetic changes in goal traits when selecting for YBW, GFW and WRS is 1.27, 0.77 and 1.33, respectively (Table 3).

Table 3: Response to selection in economically important traits from single trait selection.

Genetic response in	Economic value (\$/ewe lifetime)	Selection for YBW	Selection for GFW	Selection for WRS
CFW (kg/year)	10.7	0.02 ^a	0.07	0.02
MFD (mic/year)	-1.5	0.05	0.07	-0.10
NLW (No./year)	53.9	0.012	0	0.016
WWT (kg/year)	0.94	0.43	0.11	0.13
EBW (kg/year)	0.13	0.72	0.17	0.25
Economic response (\$/ewe lifetime)		\$ 1.27 ^b	\$ 0.77	\$ 1.33

^a Genetic response in CFW selecting for YBW:

$$R = i r_{G_{CFW} YBW} h_{CFW} h_{YBW} \sigma_{P_{CFW}} / L = 1.5 \times 0.2 \times \sqrt{0.4} \times \sqrt{0.4 \times 0.25} / 3 = 0.02 \text{ kg/year.}$$

^b Economic response in aggregate breeding value H due to selection for YBW:

$$0.2 \times 10.7 + 0.05 \times (-1.5) + 0.012 \times 53.9 + 0.43 \times 0.94 + 0.72 \times 0.13 = 1.27 \text{ \$/ewe lifetime.}$$

Since Ponzoni uses 3.85 matings per ewe lifetime, these figures would be divided by 3.85 to give increases of returns per ewe mated per year. Thus if we multiply by 1000/3.85 we would get the increase in returns per year due to genetic changes in the five breeding goal traits for a flock of 1000 breeding ewes of \$ 329.87 (1.27 × 1000/3.85) when selecting only for YBW, \$ 200.00 when selecting only for GFW and \$ 345.73 when selecting only for WRS.

Now instead of single trait selection let us select for H using a selection index combining GFW, YBW and WRS. We first construct matrices \mathbf{P} (which is symmetric) and \mathbf{G}

$$\mathbf{P} = \begin{bmatrix} 0.28 & 0.7144 & 0.1944 \\ & 20.25 & -0.4409 \\ & & 1.5 \end{bmatrix} \text{sym}$$

For example: phenotypic variance $P_{11} = \text{var}_p(1) = \sigma_p^2 = 0.28$, and phenotypic covariance of GFW and YBW $P_{12} = \text{cov}_p(1,2) = r_p \sigma_{p1} \sigma_{p2} = 0.3 \times \sqrt{0.28 \times 20.25} = 0.7144$.

$$\mathbf{G} = \begin{bmatrix} 0.0742 & 0.0768 & 0 & 0.1174 & 0.1782 \\ 0.18 & 0.4363 & 0.1055 & 3.8422 & 6.48 \\ -0.0411 & 0.2390 & -0.0385 & -0.3248 & -0.6162 \end{bmatrix}$$

For example: Genetic covariance of GFW and EBW $G_{15} = \text{cov}_G(1,5) = r_G h_1 h_5 \sigma_{p1} \sigma_{p5} = 0.2 \times \sqrt{0.35 \times 0.4 \times 0.28 \times 20.25} = 0.1782$.

The transpose of \mathbf{a} is

$$\mathbf{a}' = [10.7 \quad -1.5 \quad 53.9 \quad 0.94 \quad 0.13]$$

We then invert \mathbf{P} and get $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$ such that the transpose of \mathbf{b} is

$$\mathbf{b}' = [3.7052 \quad 0.3776 \quad -2.5429]$$

Thus the index which maximises response to selection for H is

$$I = 3.7052 \times \text{GFW} + 0.3776 \times \text{YBW} + (-2.5429) \times \text{WRS}$$

We can calculate the variance of the index $\mathbf{b}'\mathbf{G}\mathbf{a} = \mathbf{b}'\mathbf{P}\mathbf{b} = \sigma_I^2 = 15.6121$. From which $\sigma_I = 3.9512$.

Also note

$$\mathbf{P}^{-1}\mathbf{G} = \begin{bmatrix} 0.3247 & 0.1172 & 0.0061 & 0.0703 & 0.0734 \\ -0.0041 & 0.0207 & 0.0044 & 0.1835 & 0.3102 \\ -0.0707 & 0.1502 & -0.0252 & -0.1717 & -0.3291 \end{bmatrix}$$

The columns of $\mathbf{P}^{-1}\mathbf{G}$ are the sub-indexes to predict the breeding values for the traits in the objective. For example, the sub-index which predicts breeding value for trait 1 (CFW) is

$$I_1 = 0.3247 \times \text{GFW} + (-0.0041) \times \text{YBW} + (-0.0707) \times \text{WRS}$$

The efficiency of indexes with one trait dropped are $\sqrt{1 - \frac{b^2}{R_{kk}\sigma_I^2}}$, where b is the index weight of the trait dropped and R_{kk} is its diagonal element in \mathbf{P}^{-1} (not calculated above). For example,

$$\text{Efficiency when dropping GFW from index} = \sqrt{1 - \frac{3.7052^2}{4.4402 \times 15.6121}} = 0.8955.$$

Thus if we drop GFW from the full index we achieve only 89.56% of genetic gain in H . If we drop YBW from the index the efficiency is 91.52% and if we drop WRS from the index it is only 67.41%.

The genetic variance-covariance matrix \mathbf{Q} of terms in H is symmetric

$$\mathbf{Q} = \begin{bmatrix} 0.1 & 0.1212 & 0 & 0.1186 & 0.18 \\ & 2.35 & -0.0227 & 0.2299 & 0.4363 \\ & & 0.022 & 0.0445 & 0.1055 \\ & & & 2.25 & 2.9884 \\ & & & & 8.1 \end{bmatrix} \text{sym}$$

We can calculate the variance of the breeding objective $\mathbf{a}'\mathbf{Q}\mathbf{a} = \sigma_H^2 = 91.3484$. From which $\sigma_H = 9.5576$. The correlation between index and breeding objective or index accuracy is $r_{IH} = \sigma_I/\sigma_H = 0.4134$.

In comparison with single trait selection we see that with the same standardised selection differential ($i = 1.5$) and generation length ($L = 3$) the gain in economic value using the index is $\frac{1.5}{3} \sigma_I = \$ 1.98$.

Note that the best response to single trait selection was \$ 1.33 when selecting for WRS.

The covariances of economic traits with index are

$\mathbf{b}'\mathbf{G}$

$$= [3.7052 \quad 0.3776 \quad -2.5429] \begin{bmatrix} 0.0742 & 0.0768 & 0 & 0.1174 & 0.1782 \\ 0.18 & 0.4363 & 0.1055 & 3.8422 & 6.48 \\ -0.0411 & 0.2390 & -0.0385 & -0.3248 & -0.6162 \end{bmatrix}$$

$$= 0.4475 \quad -0.1584 \quad 0.1378 \quad 2.7115 \quad 4.6738$$

Multiplying by $\frac{1.5}{3} \sigma_I$ we get the expected annual gains as

$$= 0.0566 \quad -0.0200 \quad 0.0174 \quad 0.3431 \quad 0.5914$$

If these are multiplied by their economic values (a 's) we have

$$\$ 1.98 = \$ 0.61 \quad \$ 0.03 \quad \$ 0.94 \quad \$ 0.32 \quad \$ 0.08$$

Showing the relative contribution of different traits to overall gain in economic value.

Example 2: Restricted selection index construction and evaluation

This example published by Cunningham et al. (1970), illustrates the case of a selection index with one objective trait restricted to zero change. The calculations are also replicated in an Excel file¹⁰.

The restricted index method is applied to a sheep flock selected to improve fleece weight (FW), number of lambs (NL) and weaning weight (WW) but maintain ewe body weight (EBW) unchanged. The parameters assumed are in Table 4.

¹⁰ Excel file [Cunningham et al 1970 restricted selection index.xlsx](#) available as supplementary material.

Table 4: Economic value, phenotypic and genetic parameters in sheep from Cunningham et al. (1970). Phenotypic correlations above diagonal, genetic correlations below diagonal.

Parameter/trait	FW	NL	WW	EBW
Heritability	0.35	0.10	0.25	0.35
Phenotypic variance	0.25	0.25	36.0	64.0
Economic value	2.00	22.00	1.00	0.00
Fleece weight (FW)		0.00	0.00	0.00
No. of lambs (NL)	-0.10		0.00	0.00
Weaning weight (WW)	0.10	0.00		0.00
Ewe body weight (EBW)	0.20	0.30	0.20	

Using parameters from Table 4, the unrestricted index equations, including in $G\mathbf{a}$ the trait to be restricted, are

$$P\mathbf{b} = G\mathbf{a} =$$

$$\begin{bmatrix} 0.25 & 0.00 & 0.00 \\ 0.00 & 0.25 & 0.00 \\ 0.00 & 0.00 & 36.0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0.0875 & -0.0047 & 0.0887 & 0.2800 \\ -0.0047 & 0.0250 & 0.0000 & 0.2245 \\ 0.0887 & 0.0000 & 9.0000 & 2.8397 \end{bmatrix} \begin{bmatrix} 2 \\ 22 \\ 1 \\ 0 \end{bmatrix}.$$

The restricted index based on the same three variates, but which produces zero change in EBW is obtained by augmenting the equations of $P\mathbf{b}$ and give the following set of equations

$$P^*\mathbf{b}^* = G^*\mathbf{a} =$$

$$\begin{bmatrix} 0.25 & 0.00 & 0.00 & 0.2800 \\ 0.00 & 0.25 & 0.00 & 0.2245 \\ 0.00 & 0.00 & 36.0 & 2.8397 \\ 0.2800 & 0.2245 & 2.8397 & 0.0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_d \end{bmatrix} = \begin{bmatrix} 0.0875 & -0.0047 & 0.0887 & 0.2800 \\ -0.0047 & 0.0250 & 0.0000 & 0.2245 \\ 0.0887 & 0.0000 & 9.0000 & 2.8397 \end{bmatrix} \begin{bmatrix} 2 \\ 22 \\ 1 \\ 0 \end{bmatrix}.$$

Table 5 shows the resulting solutions for both indexes. The value of the variate represents the percent reduction in rate of genetic gain for aggregate genotype if that variable is dropped from the index. The value of the dummy variate is -90.03 which indicates that the expected genetic gain will increase by this percentage if the restriction is dropped. This can be also seen in the accuracy of the restricted index which resulted in $r_{IH} = 0.22$ about half of the accuracy of the unrestricted index which is $r_{IH} = 0.41$. The example illustrates the cost of restricting EBW to zero change. Cunningham mentions that this cost can be minimized if EBW would be included in the index.

Table 5: Index weightings for the regular index and the restricted index.

Variate	Unrestricted index		Restricted index	
	Index weight (\mathbf{b})	Value of variate	Index weight (\mathbf{b})	Value of variate
Fleece weight	0.643	1.44	-1.462	73.18
No. of lambs	2.163	17.76	0.475	3.95
Weaning weight	0.255	40.65	0.107	35.76
Dummy			1.880	-90.03

Note: minor differences with published results are due to rounding errors.

2.4 Computer tool to construct and evaluate unrestricted and restricted selection indexes

Calculating the index weights (\mathbf{b}) can be very tedious and several computer programs have been written to facilitate this work. For example, Cunningham & Mahon (1977) wrote the popular program SELIND.FOR for this purpose. The program is very flexible and gives index weights \mathbf{b} , the sub-indexes from columns $\mathbf{P}^{-1}\mathbf{G}$, accuracies and values of variates.

SELIND.FOR reads the run options and matrices \mathbf{P} , \mathbf{G} , \mathbf{a} and \mathbf{Q} from an input file. Preparing this input file with the necessary covariances, when these have to be calculated from heritabilities and correlations can be very tiresome, especially when different breeding objectives, selection indexes, assumed parameters and run options are to be evaluated.

For this reason a program, called [selindex.f95](#)¹¹, was written which constructs the selind input file with parameters in terms of heritabilities and correlations and runs selind as a subroutine at once. This program reads a file called [param.prn](#) with a table of up to ten traits which allows the user to select desired traits in I and in H . To make things even easier an Excel template file, [param.xlsx](#) is provided which can be filled/updated and exported as [param.prn](#). In summary, index weight calculations and index evaluation are done in two steps:

- Step 1: Set [param.xlsx](#) selecting goal traits and index traits by replacing corresponding zeros (0) with ones (1), modifying parameters if necessary and export as [param.prn](#).
- Step 2: Compile and run [selindex.f95](#) and you get the results in file [selind_out](#). You will obtain in the same folder the selind input file [selind_in](#) and a [param_check](#) file which might be useful to revise for input errors.

¹¹ Computer program [selindex.f95](#) available as supplementary material.

Trait_name	CFW	MFD	NLW	WW	EBW	GFW	YBW	WRS	XX	XX
Goal_trait	1	1	1	1	1	0	0	0	0	0
Inde_trait	0	0	0	0	0	1	1	1	0	0
Kinship	1.000	1.000	0.500	1.000	0.500	1.000	1.000	1.000	0.000	0.000
Phenot_var	0.250	4.700	0.220	9.000	20.250	0.280	20.250	1.500	0.000	0.000
Econ_value	10.700	-1.500	53.900	0.940	0.130	0.000	0.000	0.000	0.000	0.000
CFW	0.400	0.200	0.000	0.250	0.300	0.850	0.300	0.200	0.000	0.000
MFD	0.250	0.500	0.100	0.100	0.130	0.130	0.130	0.140	0.000	0.000
NLW	0.000	-0.100	0.100	0.120	0.150	0.000	0.150	-0.200	0.000	0.000
WW	0.250	0.100	0.200	0.250	0.350	0.250	0.350	-0.080	0.000	0.000
EBW	0.200	0.100	0.250	0.700	0.400	0.300	0.700	-0.080	0.000	0.000
GFW	0.750	0.160	0.000	0.250	0.200	0.350	0.300	0.300	0.000	0.000
YBW	0.200	0.100	0.250	0.900	0.800	0.200	0.400	-0.080	0.000	0.000
WRS	-0.150	0.180	-0.300	-0.250	-0.250	0.200	-0.250	0.500	0.000	0.000
XX	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
XX	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Figure 2: Excel parameter file for the example of Ponzoni (1979).

Figure 2 shows the Excel template file to be filled with the input information. In this case the parameters are those of Example 1 previously described¹². Note that only blue text should be modified. Up to 10 trait names can be defined. Goal traits and index traits are selected by replacing “0” with “1”. An additional input row named “Kinship” allows to specify the genetic relation between the animal measured and the selection candidate. For example, if NLW is in the index (specified with a “1”) measured on the dam of the selection candidate, the “Kinship” is 0.5. If a trait is measured on halve-sibs the “Kinship” would be 0.25. For traits measured on the selection candidate “Kinship” is 1.0.

Phenotypic variances for all traits have to be introduced and economic values or relative economic values for goal traits have to be set. For all traits in the objective or in the index, parameters have to be introduced with heritabilities on the diagonal, phenotypic correlations above the diagonal and genetic correlations below the diagonal. Exporting this Excel file as the space-limited text file `param.prn` and running `selindex.f95` will give the Selind output of Figure 3¹³.

¹² Excel file `param_example 1 Ponzoni.xlsx` available as supplementary material.

¹³ When using the Plato Fortran program environment make sure that your folder address is not too long. `Selindex.f95` may be unable to assign results to the output file.

```

selind_out: Bloc de notes
Archivo Edición Formato Ver Ayuda
SELECTION INDEX: MAIN INDEX

      VARIATE      B-VALUE      VALUE OF VARIATE (1)
1  GFW            3.7055            10.4491
2  YBW            0.3775             8.4775
3  WRS           -2.5431            32.5967

VARIANCE OF INDEX      15.6134
STD DEV OF INDEX      3.9514

VARIANCE OF AGG GENOT  91.3492
STD DEV OF AGG GENOT  9.5577 (2)

CORRELATION OF INDEX AND AGG GENOT  0.4134

REGRESSION OF EACH TRAIT ON INDEX (b*G/var(I)), CORRELATION OF EACH TRAIT WITH INDEX, PERCENT OF ECONOMIC GAINS (3)

      TRAIT      REGRESSION      CORRELATION      PCT OF GAIN
1  CFW          0.0287          0.3582          30.6719
2  MFD         -0.0102         -0.0262          1.5227
3  NLW          0.0088          0.2352          47.5894
4  MW           0.1737          0.4575          16.3245
5  EBW          0.2993          0.4156          3.8915

(1) VALUE OF EACH VARIATE IN THE INDEX = PERCENT REDUCTION
    IN RATE OF OVERALL GENETIC GAIN IF THAT VARIATE IS OMITTED.
(2) THIS IS THE VALUE, IN ECONOMIC UNITS, OF THE GENETIC GAIN IN AGGREGATE
    GENOTYPE ACHIEVED BY ONE STANDARD DEVIATION OF SELECTION ON THE INDEX.
(3) THESE FIGURES ARE THE PERCENTAGES OF TOTAL GAIN (2)
    ACCOUNTED FOR BY GAIN IN EACH TRAIT.

-----
VARIANCE OF SUBINDEX:  0.0263  CORRELATION OF SUBINDEX AND TRAIT:  0.5126
      VARIATE      B-VALUE      VALUE OF VARIATE (1)
1  GFW            0.3247            68.9913
2  YBW           -0.0041             0.5721
3  WRS           -0.0707            13.4253
  
```

Figure 3: Selind output extract with input parameters of Figure 2.

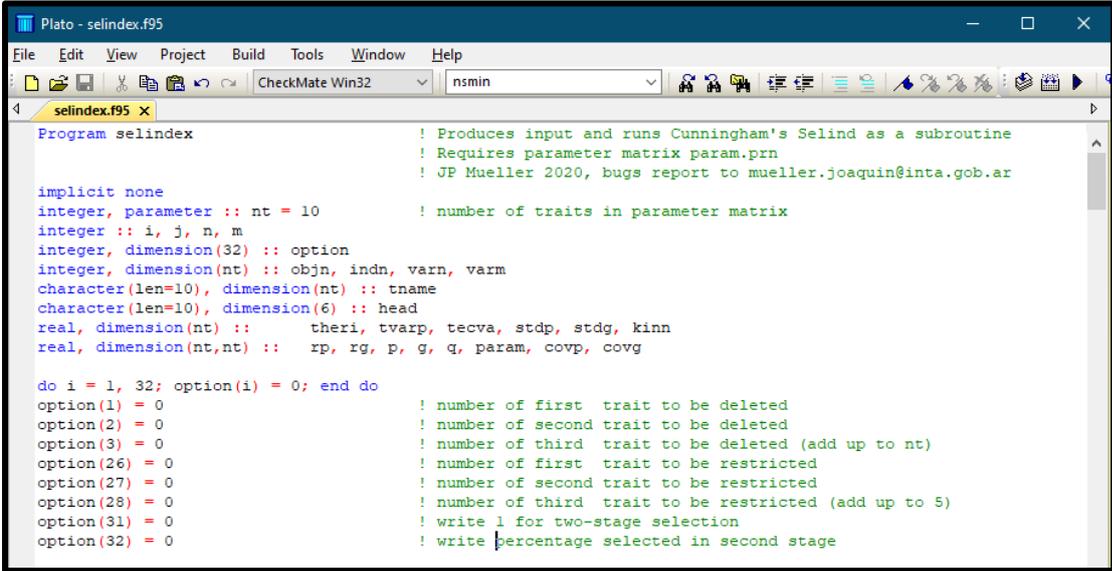
Apart from rounding, these results are the same as in the above text and also the same as in the worked Excel file ([Ponzoni 1979 selection index.xlsx](#)). Interpretation of the results is straightforward.

- The selection index I which maximises breeding objective function H is $I = 3.7055 \times GFW + 0.3775 \times YBW - 2.5431 \times WRS$.
- For example, if greasy fleece weight (GFW) is omitted in the index, overall genetic gain in monetary units will be reduced by 10.4491%.
- The accuracy of the index is $r_{IH} = 0.4134$.
- For example, genetic improvement in clean fleece weight (CFW) will be 0.0287 kg per generation and per one standard deviation of index selection differential. Thus, multiplying 0.0287 by $\sigma_I \times i/l$ gives the actual gain per year. The standard deviation of the index is 3.9514 and if $i/l = 1.5/3$, the annual genetic improvement of CFW using the index will be $0.0287 \times 3.9514 \times 0.5 = 0.0567 \text{ kg/year}$.
- For example if we want to know the sub-index which predicts breeding value for CFW we should use $I_{CFW} = 0.3247 \times GFW - 0.0041 \times YBW - 0.0707 \times WRS$.

Cunningham's Selind program also allows to compare the effects of deleting one or more traits, restricting to no change one or more traits and selecting in two stages given a proportion selected in the first stage. These options can be easily set in [selindex.f95](#) by specifying the trait numbers to be deleted, restricted or selected in stages.

Figure 4 shows the initial statements of the program with the corresponding options. For example to reproduce Cunningham's (1970) example with ewe body weight (trait number 4) restricted to zero genetic change, set **option(26) = 4** in [selindex.f95](#) and run with [param.prn](#)

file obtained from [param_example 2 Cunningham.xlsx](#)¹⁴ and the results of the paper or of the worked example will be obtained apart from rounding errors.



```

Plato - selindex.f95
File Edit View Project Build Tools Window Help
selindex.f95 x
Program selindex           ! Produces input and runs Cunningham's Selind as a subroutine
                          ! Requires parameter matrix param.prn
                          ! JP Mueller 2020, bugs report to mueller.joaquin@inta.gob.ar

implicit none
integer, parameter :: nt = 10           ! number of traits in parameter matrix
integer :: i, j, n, m
integer, dimension(32) :: option
integer, dimension(nt) :: objn, indn, varn, varm
character(len=10), dimension(nt) :: tname
character(len=10), dimension(6) :: head
real, dimension(nt) ::   theri, tvarp, tecva, stdp, stdg, kinn
real, dimension(nt,nt) ::  xp, rg, p, g, q, param, covp, covg

do i = 1, 32: option(i) = 0; end do
option(1) = 0           ! number of first trait to be deleted
option(2) = 0           ! number of second trait to be deleted
option(3) = 0           ! number of third trait to be deleted (add up to nt)
option(26) = 0          ! number of first trait to be restricted
option(27) = 0          ! number of second trait to be restricted
option(28) = 0          ! number of third trait to be restricted (add up to 5)
option(31) = 0          ! write 1 for two-stage selection
option(32) = 0          ! write percentage selected in second stage

```

Figure 4: Capture of initial `selindex.f95` statements showing run options for traits to be deleted, restricted or selected in stages.

¹⁴ Excel file [param_example 2 Cunningham.xlsx](#) available as supplementary material.

3. Estimation of breeding values in populations with uncertain sires

3.1 Introduction

With multiple-sire joining, sometimes the sire of a progeny cannot be assigned with certainty and best linear unbiased prediction (BLUP) breeding values are calculated setting the sire to “unknown” or treating the possible sires as a particular mating-group. In both cases breeding values will be calculated with larger prediction error variances (PEV) and serious underestimation of genetic trend compared to situations when sires are known (Scarth and Graser, 1987). When possible sires can be identified, the sire of a progeny is “uncertain” rather than “unknown” and this can be accounted for in the relationship matrix (**A**) necessary to calculate breeding values. It has been shown that identifying possible sires of a progeny and assigning to each a mating probability, BLUP breeding values can be predicted with lower PEV compared to situations with unknown sires or mating groups and genetic trend can be estimated almost without bias (Perez-Enciso and Fernando, 1992; Sullivan, 1995).

The present chapter describes a procedure to process field information of populations with uncertain sires. The procedure is based on the application of a sequence of computer programmes which process field data and pedigree with possible sires to produce the necessary input files and run the publicly available **wombat**¹⁵ software (Meyer, 2007) to predict BLUP breeding values with typical statistical models.

3.2 Uncertain sire pedigree: theoretical background

Standard BLUP breeding value prediction with the animal model requires the relationship matrix **A** and its inversion. With single sire mating Henderson (1976) presented a convenient method to set up **A**⁻¹ without actually computing **A**. the method to obtain **A**⁻¹ is as follows:

- Write a matrix **L** from pedigree such that $\mathbf{LL}' = \mathbf{A}$.
- **L** can be written as **TD** where **D** is a diagonal matrix with diagonals equal to the diagonals of **L**, thus $\mathbf{T} = \mathbf{LD}^{-1}$.
- Then $\mathbf{A}^{-1} = (\mathbf{LL}')^{-1} = (\mathbf{TD}^2\mathbf{T}')^{-1} = (\mathbf{T}')^{-1}\mathbf{D}^{-2}\mathbf{T}^{-1}$.

Since **T**⁻¹ and **D** can be easily constructed we get **A**⁻¹ from rather simple calculations. This method to obtain **A**⁻¹ is widely used in BLUP breeding value prediction software such as **wombat**. With uncertain sires the relationship matrix **A** needs to consider the mating probabilities of the possible sires. Henderson (1988) adjusted the previous procedure to obtain a “uncertain sire” **A**⁻¹ by computing **D** and writing **T**⁻¹ from pedigree and sire probabilities. He describes a tabular method to construct **A**⁻¹ manually using a very small pedigree as an example. While **T**⁻¹ has an extremely simple form, **D** is rather laborious to construct in this way and for large pedigrees a convenient computer program becomes necessary.

¹⁵ Breeding value estimation software “**wombat**” available at <http://didgeridoo.une.edu.au/km/wmbdownload1.php>

3.3 Overview of the procedure

Suppose we have individual animal field information including measurements (for example 6-month weight, dam's average daily milk yield, etc.), fixed effects (for example birth year, gender, etc.) and genealogy (dam and sire) digitalized in a database such as an Excel spreadsheet. The dam or the sire of an animal may be unknown or the sire is uncertain and possible sires are known. With uncertain sires the pedigree file will include the possible sires and their mating probability. For example, an individual may have two possible sires, A and B with mating probabilities p_A and p_B such that $p_A + p_B = 1.0$. If we have no reason to expect a different number of progeny from each sire, mating probabilities would be $p_A = p_B = 0.5$. But suppose sire A remained two months in service and sire B only one month, then we could speculate mating probabilities to be $p_A = 0.75$ and $p_B = 0.25$. Another example for assigning different mating probabilities could be if sires are of different age. We could speculate that young sires mate fewer females than grown sires and guess suitable mating probabilities for each possible sire. Here we assume equal mating probabilities for possible sires. Minor adjustments to the procedures would be needed to allow for unequal probabilities.

The procedure to predict breeding values from field data and uncertain pedigree involves several steps. Firstly, the animals of interest need to be extracted from the raw database. For example, those animals with records for the trait/s studied or those with at least one known parent or with possible sires. Secondly, animal tags have to be renumbered sequentially to facilitate further processing. Thirdly, the uncertain sire relationship matrix \mathbf{A} has to be constructed. More precisely, the non-zero elements of \mathbf{A}^{-1} are needed. Finally, wombat has to be run with the appropriate input and parameter files. These three steps are best aided by computer programmes which read field data from spreadsheet databases and produce the required wombat input files.

Field information can include different traits and wombat analyses can be done with different statistical models. Thus, specific computer programmes are needed to handle each case. This guideline is based on field information and wombat input files common for breeding value estimation in sheep and goat community-based breeding programmes (CBBP) and the necessary computer programmes were developed to target such cases. The computer programs are provided as supplementary material and can be used without computer programming knowledge. Basic knowledge of Fortran programming is however necessary to modify the programs if different field information or wombat input files are chosen.

3.4 Work flow

Let us suppose we have raw field data in an Excel file which we export as a space-separated text file called `main.prn`. First we extract from this file the animals of interest with programme `main.f95`. Second we renumber the pedigree and data files with programs `renped.f95` and `rendat.f95`, respectively. Third we obtain the non-zero elements of \mathbf{A}^{-1} with programme `uncert.f95`. The resulting output files will be `dat.prn` and `animal.gin` which are the input files to run wombat and predict breeding values for populations with uncertain sires. The programmes need to be executed in sequence. The mandatory work flow is outlined in Figure 5.

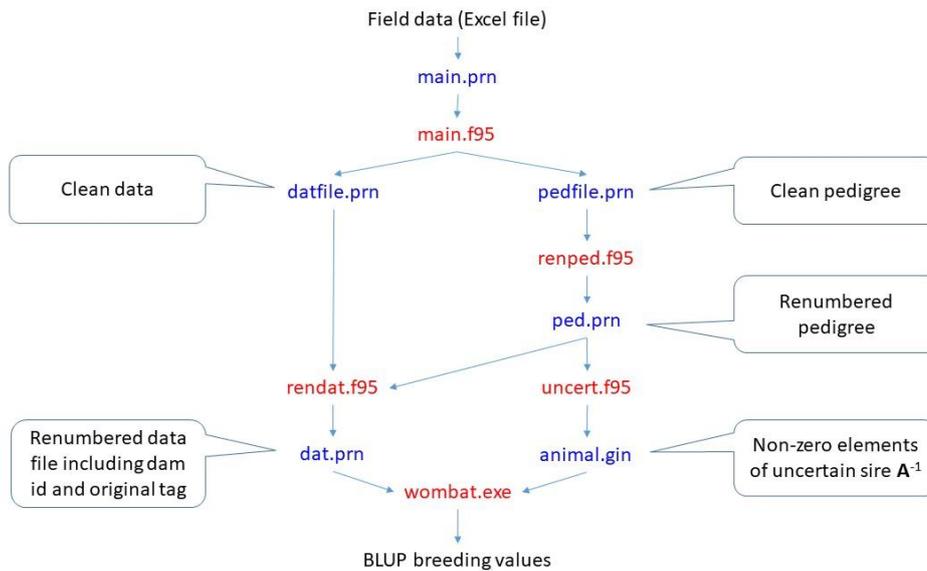


Figure 5: Work flow to obtain BLUP breeding values from field data of populations with uncertain sires. In blue font are input/output files and in red font computer programs.

3.5 Example of estimation of breeding values with uncertain sires

The procedure is illustrated with an example of a goat population with incomplete pedigree including some uncertain sires, various fixed effects and live weight records at different ages. The field data are supposed to be in an Excel file¹⁶ (Figure 6).

seq_animal	seq_dam	seq_sire1	seq_sire2	seq_sire3	animal_ori	village	month	season	year	sex	bt	ewppwt	BWT	WWT	SMWT	NNWT	YWT
1148	0	0	0	0	20045	2	6	2	2018	2	1	29	2.2				
1149	0	0	0	0	20678	2	11	1	2014	2	1	23	1.4	4.8	14	14	16
1150	0	0	0	0	20682	2	12	1	2014	2	1	25.2	2		13.2	12.8	16
1151	0	0	0	0	20685	2	12	1	2014	2	1	22	1.6	6.4	10.8	12	16
1152	0	0	0	0	21515	2	12	1	2016	1	1	27.6	2.4				18.8
1153	0	0	0	0	21516	2	12	1	2016	1	1	25.4	2				
1154	0	0	0	0	21517	2	12	1	2016	1	1	28.4	2.4	7	6.4		18.8
1155	0	0	0	0	21620	2	4	2	2017	1	1	23	3				
1156	0	0	0	0	21701	2	4	2	2017	1	1	23	2.8				
1157	0	0	0	0	21703	2	8	1	2014	1	1	19.2	2.4				
1158	0	0	0	0	21704	2	5	2	2017	1	1	25	2.8				
1159	0	0	0	0	21706	2	5	2	2017	2	1	20	2.2				
1160	0	0	0	0	21707	2	5	2	2017	1	1	22.8	2.4				
1161	0	0	0	0	22188	2	12	1	2016	1	1	32	2.6	9			
1162	0	0	0	0	22189	2	12	1	2016	1	1	31	2.4	7.8	10.2		
1163	0	0	0	0	32661	1	12	1	2017	1	1	2					
1164	0	0	0	0	32700	1	12	1	2017	1	1	2		7			
1165	153	0	0	0	21692	2	9	1	2016	2	1	28.4	2.4				
1166	0	0	0	0	30355	1	12	1	2015	2	1	2					
1167	0	0	0	0	30735	1	10	1	2017	1	1	2.4	6.2	7			
1168	515	0	0	0	31044	1	11	1	2013	1	1	2.4	9	10	12.8	14	
1169	523	0	0	0	31045	1	11	1	2013	1	1	2.2	11	11	13.6	14.2	
1170	555	0	0	0	31197	1	11	1	2013	2	1	20.2	2.5	7	8	10	12.2
1171	0	1170	0	0	31325	1	10	1	2015	2	1	1.75					
1172	0	1170	0	0	31326	1	12	1	2015	1	1	2		10			
1173	0	1170	0	0	31328	1	12	1	2015	1	1	2					
1174	0	1170	0	0	31332	1	12	1	2015	2	1	2					
1175	934	1168	0	0	31487	1	12	1	2014	1	1	2.27	6.4	7.4	12	15.4	
1176	0	0	0	0	32393	1	10	1	2017	1	1	2	7	7.4			
1177	116	0	0	0	20578	2	8	1	2013	2	1	24	1.8				
1178	286	0	0	0	20658	2	9	1	2013	2	1	28.2	2	3.6	9	16	16.4
1179	0	0	0	0	21386	2	11	1	2017	2	1	25.2	2	6.8	8		15
1180	483	0	0	0	21387	2	11	1	2017	2	1	25.6	2	6.4	9	16.8	12.6
1181	415	42	43	0	21406	2	11	1	2015	1	1	26	2.4	8	7.2	12.2	
1182	0	28	1165	0	21511	2	3	2	2014	1	1	24.6	2	5	13.8		20

Figure 6: Initial rows of example Excel file with typical field data and with up to three possible sires.

¹⁶ Excel file [main_data_seq.csv](#) available as supplementary material.

A column header row is expected in this Excel file. Any column heading text is accepted but will not be used during processing. The first twelve columns must contain integer values for the pedigree and coded fixed or random effects whereas the following six columns must contain real values for measurements. Thus, only numbers are admitted in this file. The maximum number of digits per column is eight (or ten), if there are more digits they will be left-truncated. Any number can be used to identify animals, dams and sires as long as an animal number is larger than any of its parents' number. Missing dam or missing sires must be set to 0 (zero). Export this Excel file as a space-separated free-format text file named [main08.prn](#) (or [main10.prn](#) if you have ten digits per column) and run sequentially programs [main08.f95](#) (or [main10.f95](#) if you use [main10.prn](#)), [renped.f95](#), [rendat.f95](#) and [uncert.f95](#)¹⁷. Note that the first program ([main.f95](#)) will read [main.prn](#) expecting the following information per column regardless of the original column heading:

- Animal id
- Dam id
- Sire 1 id (known sire or first possible sire)
- Sire 2 id (second possible sire)
- Sire 3 id (third possible sire)
- Original id (any number)
- Fixed or random effect 1 (in the example: village)
- Fixed or random effect 2 (in the example: birth month)
- Fixed or random effect 3 (in the example: season)
- Fixed or random effect 4 (in the example: birth year)
- Fixed or random effect 5 (in the example: gender)
- Fixed or random effect 6 (in the example: birth type)
- Trait as co-variable (in the example: dam post-partum weight)
- Trait 1 (in the example: birth weight)
- Trait 2 (in the example: weaning weight)
- Trait 3 (in the example: 6-month weight)
- Trait 4 (in the example: 9-month weight)
- Trait 5 (in the example: yearling weight)

After having run the four programmes in sequence, five output files will have been created: the purged pedigree and data files [pedfile.prn](#) and [datfile.prn](#), the renumbered pedigree and data files [ped.prn](#) and [dat.prn](#) and a file called [animal.gin](#) which contains the non-zero elements of the inverse “uncertain sire” relationship matrix. The output files [dat.prn](#) [animal.gin](#) are the required input files to run [wombat](#) with the corresponding parameter file.

¹⁷ Computer programs [main08.f95](#), [main10.f95](#), [renped.f95](#), [rendat.f95](#) and [uncert.f95](#) available as supplementary material.

Notes regarding the “uncertain sire” programs

The programs are set to work with a maximum of 10000 animals, 6 fixed or random effects, 5 traits and 1 co-variable trait. If there are more animals (rows) in the Excel file (in the example: Main_Data_seq.csv) the maximum accepted by the program has to be increased manually in the four programmes: **main08.f95** (**main10.f95**), **renped.f95**, **rendat.f95** and **uncert.f95**. This is an easily recognisable task to do in the initial rows of the programs and subroutines. Also, more than 6 fixed effects and/or more than 5 traits can be programmed. For example, if a 7th effect (eff7) and/or a 6th trait (tr6) is needed add “eff7” and/or “tr6” to the dimension, read and write statements and add one element in the read and write format statements in programs **main08.f95** (**main10.f95**) and **rendat.f95**.

How to organize your computer to run the “uncertain sire” programs

Copy the following files into a folder:

- **main.prn**: this is the space-separated exported field Excel input file
- **main.f95**: this program purges the input file and produces **pedfile.prn** and **datfile.prn**
- **renped.f95**: this program renumbers ids in pedfile and produces **ped.prn**
- **rendat.f95**: this program renumbers ids in datfile and produces **dat.prn**
- **uncert.f95**: this program generates the uncertain sire \mathbf{A}^{-1} and produces **animal.gin**

The Fortran programs have to be compiled to obtain the executables. This is best done with the program **Plato**¹⁸. If you have not linked the path to **Plato** and **wombat** you may also need to include these two programs into the same folder. You may also create and copy into this folder the wombat parameter file (wombat.par). Thus, running in sequence all the above programs including wombat, you will obtain the uncertain sire breeding values for your animals in the Excel file.

¹⁸ Fortran 95 software “**Plato**” available for personal use at <https://www.silverfrost.com/>

Supplementary material:

The following supplementary materials are available:

Chapter 1: Breeding objectives

- [Ponzoni 1986 breeding objective.xlsx](#)
- [Mueller et al 2021 breeding objective.xlsx](#)
- [Amer 1999 DGE.xlsx](#)
- [AmerDGE.for](#)
- [inAmer](#)
- [outAmer](#)

Chapter 2: Selection indexes

- [Ponzoni 1979 selection index.xlsx](#)
- [Cunningham et al 1970 restricted selection index.xlsx](#)
- [Walsh and Lynch 2000 desired gain selection index.xlsx](#)
- [selindex.f95](#)
- [param_example 1 Ponzoni.xlsx](#)
- [param_example 2 Cunningham.xlsx](#)

Chapter 3: Uncertain sires

- [main_data_seq.csv](#)
- [main08.f95](#)
- [main10.f95](#)
- [renped.f95](#)
- [rendat.f95](#)
- [uncert.f95](#)

References

- Amer, P.R. 1999. Economic accounting of numbers of expressions and delays in sheep genetic improvement. *New Zealand Journal of Agricultural Research* 42, 325-336.
- Cunningham, E.P. and Mahon, G.A.T. 1977. SELIND. Users' Guide. A FORTRAN computer program for genetic selection indexes. An Foras Taluntais. Dunsinea, Ireland.
- Cunningham, E.P., Moen, R.A. and Gjedrem, T. 1970. Restriction of selection indexes. *Biometrics*, 26, 67-74.
- Henderson, C.R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32, 69-83.
- Henderson, C.R. 1988. Use of an average numerator relationship matrix for multiple sire joining. *Journal of Animal Science* 66, 1614-1621.
- James, J.W. 1982. Construction, uses and problems of multitrait selection indices. *2nd World Congress on Genetics Applied to Livestock Production* 5, 130-139.
- Kargar Borzi, N., Ayatollahi Mehrgardi, A. and Abassi, M.A. 2017. *Iranian Journal of Applied Animal Science* 7, 631-636.
- Kempthorne, O., Nordskog, A.W. 1959. Restricted selection indexes. *Biometrics* 15, 10-19.
- Kosgey, I.S., van Arendonk, J.A.M. and Baker, R.L. 2003. Economic values for traits of meat sheep in medium to high production potential areas of the tropics. *Small Ruminant Research* 50, 187-202.
- McClintock, A.E. and Cunningham, E.P. 1974. Selection in dual purpose cattle populations: defining the breeding objective. *Animal Production* 18, 237-247.
- Meyer, K. (2007). WOMBAT – A tool for mixed model analyses in quantitative genetics by REML. *Journal of Zhejiang University Science B* 8, 815-821.
- Mueller, J.P., Getachew, T., Rekik, M., Rischkowsky, B., Abate, Z., Goshme, S., Wale, Y. and Haile, A. 2021. Converting multi-trait breeding objectives into operative selection indexes to ensure genetic gains in low input sheep and goat breeding programs. *Animal (accepted)*.
- Mrode, R.A. 2005. *Linear models for the prediction of animal breeding values*, 2nd edition. CABI Publishing, p. 20.
- Perez-Enciso, M. and Fernando, R.L. 1992. Genetic evaluation with uncertain parentage: a comparison of methods. *Theoretical and Applied Genetics* 84, 173-179.
- Pesek J. and Baker R.J. 1969. Desired improvement to selection indexes. *Canadian Journal of Plant Science* 49, 803-804.
- Ponzoni, R.W. 1979. Objectives and selection criteria for Australian Merino sheep. *Proceedings of the Australian Association of Animal Breeding and Genetics* 1, 320-336.

- Ponzoni, R.W. 1982. Breeding objectives in sheep improvement programmes. *2nd World Congress on Genetics Applied to Livestock Production 5*, 619-634.
- Ponzoni, R.W. 1986. A profit equation for the definition of the breeding objective of Australian merino sheep. *Journal of Animal Breeding and Genetics 103*, 342-357.
- Ponzoni, R.W. 1988. The derivation of economic values combining income and expense in different ways: An example with Australian Merino Sheep. *Journal of Animal Breeding and Genetics 105*, 143-153.
- Scarth, R.D. and Graser, H.-U. 1987. Comparisons of genetic evaluation from animal models for single and multiple sire joinings. *Proceedings of the Australian Association of Animal Breeding and Genetics 6*, 369-372. <http://www.aaabg.org/proceedings/1987/ab87075.pdf>
- Sullivan, P.G. 1995. Alternatives for genetic evaluation with uncertain parentage. *Canadian Journal of Animal Science 75*, 31-36. <https://cdnsciencepub.com/doi/pdf/10.4141/cjas95-004>
- Walsh, B. and Lynch, M. 2000. Theory of index selection (draft). *Evolution and Selection of Quantitative Traits*. Oxford University Press, p. 409-441. <http://nitro.biosci.arizona.edu/zdownload/Volume2/Chapter23.pdf>