

Research

Open Access

Fuzzy classification of phantom parent groups in an animal model

Freddy Fikse

Address: Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 75007 Uppsala, Sweden

Email: Freddy Fikse - Freddy.Fikse@hgen.slu.se

Published: 28 September 2009

Received: 20 August 2009

Genetics Selection Evolution 2009, **41**:42 doi:10.1186/1297-9686-41-42

Accepted: 28 September 2009

This article is available from: <http://www.gsejournal.org/content/41/1/42>

© 2009 Fikse; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Genetic evaluation models often include genetic groups to account for unequal genetic level of animals with unknown parentage. The definition of phantom parent groups usually includes a time component (e.g. years). Combining several time periods to ensure sufficiently large groups may create problems since all phantom parents in a group are considered contemporaries.

Methods: To avoid the downside of such distinct classification, a fuzzy logic approach is suggested. A phantom parent can be assigned to several genetic groups, with proportions between zero and one that sum to one. Rules were presented for assigning coefficients to the inverse of the relationship matrix for fuzzy-classified genetic groups. This approach was illustrated with simulated data from ten generations of mass selection. Observations and pedigree records were randomly deleted. Phantom parent groups were defined on the basis of gender and generation number. In one scenario, uncertainty about generation of birth was simulated for some animals with unknown parents. In the distinct classification, one of the two possible generations of birth was randomly chosen to assign phantom parents to genetic groups for animals with simulated uncertainty, whereas the phantom parents were assigned to both possible genetic groups in the fuzzy classification.

Results: The empirical prediction error variance (PEV) was somewhat lower for fuzzy-classified genetic groups. The ranking of animals with unknown parents was more correct and less variable across replicates in comparison with distinct genetic groups. In another scenario, each phantom parent was assigned to three groups, one pertaining to its gender, and two pertaining to the first and last generation, with proportion depending on the (true) generation of birth. Due to the lower number of groups, the empirical PEV of breeding values was smaller when genetic groups were fuzzy-classified.

Conclusion: Fuzzy-classification provides the potential to describe the genetic level of unknown parents in a more parsimonious and structured manner, and thereby increases the precision of predicted breeding values.

Background

Historically, genetic groups have been included in genetic evaluation models to account for selection not described by known genetic relationships. Introduction of animal models, which makes it possible to account for known

relationships in the genetic evaluation, reduced the need for genetic groups [1,2]. However, in practice, genetic evaluations can be hampered by incomplete pedigrees, due to, for example, deficiencies in pedigree recording systems and importation of animals.

An animal whose parent(s) are unknown can be assigned so-called phantom parents. These phantom parents are assumed to be unrelated, non-inbred and to have a single descendant. Phantom parents themselves are not of interest, but are considered only to facilitate modelling and computations [3].

The strategy for assigning unknown parents to genetic groups should reflect the average genetic level of unknown parents [1]. Differences in the genetic level between sub-populations of unknown parents are a good reason to form genetic groups for each sub-population, thereby avoiding the assumption that base animals belong to a single population. Common factors considered in the definition of genetic groups for unknown parents are birth year of progeny, selection intensity (selection path) and origin [3-5].

Except for a few examples where genetic groups are clearly distinct, definitions of genetic groups are often based on arbitrary rules. An accurate modelling of the expected breeding values of unknown parents will lead to the creation of many groups, each one with only a few animals. However, the drawbacks of such a strategy are either confounding with other fixed effects in the model [6] or imprecise solutions for genetic group effects. Thus, definition of genetic groups should not be too precise to yield sufficiently large groups. Moreover, if incorrect information about a base animal's attributes (e.g., birth year and origin) is used to assign the unknown parents to a group, then genetic groups will not reflect the expected genetic merit of the unknown parents. These aspects of uncertainty and inaccuracy are usually not adequately handled in the allocation of unknown parents to genetic groups.

To improve the adjustment for seasonal effects in the genetic evaluation, Strandberg and Grandinson [7] have suggested to assign a cow's record partially to the herd-year-season of calving, and partially to the closest adjoining class. They have labelled the approach "fuzzy classification", after fuzzy logics, a methodology used in expert systems to handle inexact reasoning.

The aim of this study was to develop an algorithm for fuzzy classification of phantom parent groups in an animal model. This approach is illustrated with a simulation, where fuzzy-classified genetic groups are used to handle uncertainty about the time of birth. In addition, the possibility to model a linear time trend in the average genetic level of phantom parents with a small number of genetic groups will be illustrated.

Methods

Genetic model

The infinitesimal model assumes that an animal's breeding value is the sum of the parent average and a Mendelian sampling deviation that resembles the random process of sampling parental genotypes (e.g. [8]):

$$u_i = 0.5u_s + 0.5u_d + \phi_i,$$

where u_i and N_i are the additive genetic value and Mendelian sampling deviation for animal i , respectively, and s and d are the sire and dam of animal i . In matrix representation:

$$\mathbf{u} = [\mathbf{P}_b \quad \mathbf{P}] \begin{bmatrix} \mathbf{u}_b \\ \mathbf{u} \end{bmatrix} + \boldsymbol{\phi},$$

where:

\mathbf{u}_b = vector of additive genetic effects for phantom parents;

\mathbf{u} = vector of additive genetic effects for known animals;

$[\mathbf{P}_b \quad \mathbf{P}]$ = matrix that relates parents to progeny; each row contains two non-zero elements (0.5) in the columns pertaining to the sire and dam;

$\boldsymbol{\phi}$ = vector with Mendelian sampling deviations.

Rearranging this yields the following result [6]:

$$\mathbf{u} = (\mathbf{I} - \mathbf{P})^{-1} [\mathbf{P}_b \mathbf{u}_b + \boldsymbol{\phi}].$$

That is, the additive genetic merit of all animals can be written as a linear function of the additive genetic effects of phantom parents and Mendelian sampling deviations.

Genetic groups

Inclusion of genetic groups in the model for genetic evaluation can accommodate for non-zero base population means:

$$E(\mathbf{u}_b) = \mathbf{Q}_b \mathbf{g},$$

where:

\mathbf{Q}_b = incidence matrix relating phantom parents to their respective base population means;

\mathbf{g} = vector with base population means.

In the approach by Robinson [5] and Quaas [6], the matrix \mathbf{Q}_b contains one non-zero element in each row, in

the column corresponding to the genetic group to which the phantom parent belongs. For the fuzzy classification it is proposed that any row of matrix Q_b has all elements equal to 0, except for one, two or more non-zero coefficients (elements defined by $0 \leq Q_{bij} \leq 1$) such that they add up to 1. For example, if the birth year of a base animal is estimated to be γr , the phantom parent of this animal can be allocated to genetic groups for birth year $\gamma r-1$, γr and $\gamma r+1$ with proportions 0.2, 0.6 and 0.2. This way it is possible to accommodate for uncertainty about the attributes of the base animal that are used for allocation of its phantom parents to genetic groups.

As a consequence of the non-zero base population means, the expectation of the vector with additive genetic effects becomes [6]:

$$E(\mathbf{u}) = \mathbf{Qg}$$

where:

$$\mathbf{Q} = (\mathbf{I} - \mathbf{P})^{-1} \mathbf{P}_b \mathbf{Q}_b$$

Matrix $(\mathbf{I} - \mathbf{P})^{-1} \mathbf{P}_b$ relates the breeding values of animals to phantom parents. A single non-zero element in a row represents the expected fraction of the i^{th} animal's genes derived from the j^{th} phantom parent, and the rows of this matrix sum to one [6]. Post-multiplication of this matrix with the "fuzzy" Q_b yields a matrix the rows of which sum to one, for both the distinct and fuzzy classification of genetic groups. Element q_{ij} of Q is the expected fraction of the i^{th} animal's genes deriving from the j^{th} base population. Matrix Q can be computed recursively from a list of sires and dams. Element $q_{ij} = 0.5(S + D)$, where S (or D) is q_{sj} (or q_{dj}), or the proportion of phantom parent of animal i assigned to group j .

Example

Consider the example pedigree in Figure 1. Upper case letters refer to identified animals, and lower case letters to their unknown base parents. The matrices P_b , P , (relating parents to progeny) and $(\mathbf{I} - \mathbf{P})^{-1} \mathbf{P}_b$ (relating animals to phantom parents) for this example pedigree are:

$$[P_b | P] = \begin{array}{c|cccccccc|cccccc} & a & b & c & d & e & f & g & A & B & C & D & E & F \\ \hline A & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ B & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ C & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ D & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ E & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 \\ F & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 \end{array}$$

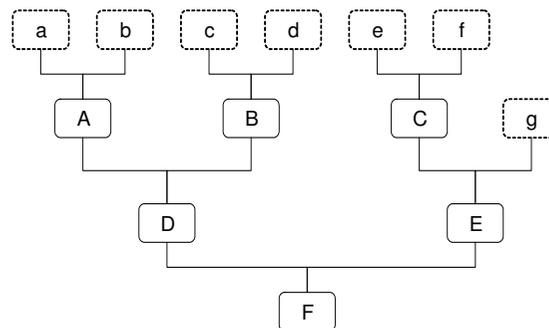


Figure 1
Pedigree example.

$$(\mathbf{I} - \mathbf{P})^{-1} \mathbf{P}_b = \begin{array}{c|cccccc} & a & b & c & d & e & f & g \\ \hline A & 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ B & 0 & 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ C & 0 & 0 & 0 & 0 & 0.5 & 0.5 & 0 \\ D & 0.25 & 0.25 & 0.25 & 0.25 & 0 & 0 & 0 \\ E & 0 & 0 & 0 & 0 & 0.25 & 0.25 & 0.5 \\ F & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.125 & 0.25 \end{array}$$

Note that matrix $(\mathbf{I} - \mathbf{P})^{-1} \mathbf{P}_b$ represents a block of matrix T that would arise from the TDT' decomposition (where D is a diagonal matrix and T a lower triangular matrix; [9]) of the relationship matrix for unknown parents a-g and know animals A-F.

Suppose that phantom parents can be related to their (three) base groups according to the following matrix:

$$Q_b = \begin{array}{ccc|c} & g1 & g2 & g3 \\ \hline 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 1 & 0 & 0 & c \\ 0 & 0.6 & 0.4 & d \\ 0 & 0 & 1 & e \\ 0 & 0.3 & 0.7 & f \\ 0.2 & 0.4 & 0.4 & g \end{array}$$

Here, phantom parents a and c came from one group, phantom parent b came from another group, and phantom parent e from a third group. Phantom parents d and f were partly assigned to the second group and the remainder to a third group, and phantom parent f was assigned to all groups, with proportions summing to one. Consequently,

$$\mathbf{Q} = \begin{matrix} & \begin{matrix} g1 & g2 & g3 \end{matrix} \\ \begin{matrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.3 & 0.2 \\ 0 & 0.15 & 0.85 \\ 0.5 & 0.4 & 0.1 \\ 0.1 & 0.275 & 0.625 \\ 0.3 & 0.3375 & 0.3625 \end{matrix} & \begin{matrix} A \\ B \\ C \\ D \\ E \\ F \end{matrix} \end{matrix}$$

Observe that the rows of matrix \mathbf{Q} sum to one. In the original specification with distinct genetic groups, matrix \mathbf{Q} would contain sums of powers of 0.5 as elements. In case of fuzzy- classified genetic groups, \mathbf{Q} can essentially contain any value between 0 and 1, depending on the fractions in \mathbf{Q}_b .

Mixed model equations

The transformed mixed model equations for an animal model with genetic groups derived by Quaas [6] are also applicable in the case of fuzzy classification of genetic groups, because the derivation is for any arbitrary matrix \mathbf{Q} . The part on the left hand side is due to relationships:

$$\mathbf{A}^* = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{Q} \\ \mathbf{Q}'\mathbf{A}^{-1} & \mathbf{Q}'\mathbf{A}^{-1}\mathbf{Q} \end{bmatrix},$$

where \mathbf{A}^{-1} is the additive genetic relationships between known animals and \mathbf{Q} , as defined before, relates breeding values of known animals to genetic group effects [6].

Considering that $\mathbf{A}^{-1} = (\mathbf{I} - \mathbf{P}')\mathbf{D}^{-1}(\mathbf{I} - \mathbf{P})$ [9] and drawing on the definition of \mathbf{Q} in 2.2, Quaas [6] showed that \mathbf{A}^* for a model with genetic groups can be written as:

$$\mathbf{A}^* = \mathbf{H}'\mathbf{D}^{-1}\mathbf{H} = \sum_i \mathbf{h}_i \mathbf{h}_i' d_i^{-1},$$

where \mathbf{h}_i' is the i^{th} row of \mathbf{H} , d_i is the i^{th} diagonal element of \mathbf{D} , $\mathbf{H} = [(\mathbf{I}-\mathbf{P}) - \mathbf{P}_b\mathbf{Q}_b]$, and $\mathbf{D}^{-1} = \text{var}(\phi)$. For animals with known parents, each row of \mathbf{H} contains three non-zero elements: a one and two negative halves. For base animals with phantom parents, each row of \mathbf{H} usually contains three or more non-zero elements: a 1 and several negative values corresponding to the phantom parents' groups. These elements have a value of $-0.5p_{ij}$, where p_{ij} is the proportion of phantom parents of i assigned to group j . This structure leads directly to an algorithm for forming \mathbf{A}^* that is a simple extension of Henderson's \mathbf{A}^{-1} algorithm. Let s_i (d_k) be the number of groups the phantom sire (dam) of i is assigned to, or 1 if the sire (dam) of i is known; m_j (f_k) the equation number of the animal's sire

(dam) or its group if the sire (dam) is unknown, and $x = d_i^{-1}$, then the following contributions need to be added: see table 1.

The rules for creating \mathbf{A}^* resemble those for forming the inverse of an average numerator relationship matrix in case there are a finite number of potential parents (e.g., [10,11] and [12]).

Example: creating \mathbf{A}^*

For the example pedigree (Figure 1), the contributions for animals A-D to elements of \mathbf{A}^* are in Table 2. For animals A, D and F there are 9 contributions to \mathbf{A}^* , for animal B and C 16,

and for animal E 25. For animal C there are 16 contributions, but only 9 elements of \mathbf{A}^* are

affected because the phantom parents of animal C are (in part) assigned to the same group. Observe that the values added to \mathbf{A}^* can become very small if phantom parents are assigned to genetic groups with low proportions, for example 0.04 to element (g_3, g_3) for animal B.

The complete \mathbf{A}^* for this example is:

$$\mathbf{A}^* = \begin{bmatrix} 1.5 & 0.5 & 0 & -1 & 0 & 0 & -0.5 & -0.5 & 0 \\ 0.5 & 1.5 & 0 & -1 & 0 & 0 & -0.5 & -0.3 & -0.2 \\ 0 & 0 & 1.333 & 0 & -0.6667 & 0 & 0.0667 & -0.0167 & -0.7167 \\ -1 & -1 & 0 & 2.5 & 0.5 & -1 & 0 & 0 & 0 \\ 0 & 0 & -0.6667 & 0.5 & 1.8333 & -1 & -0.1333 & -0.2667 & -0.2667 \\ 0 & 0 & 0 & -1 & -1 & 2 & 0 & 0 & 0 \\ -0.5 & -0.5 & 0.0667 & 0 & -0.1333 & 0 & 0.5133 & 0.4267 & 0.1267 \\ -0.5 & -0.3 & -0.0167 & 0 & -0.2667 & 0 & 0.4267 & 0.4158 & 0.2408 \\ 0 & -0.2 & -0.7167 & 0 & -0.2667 & 0 & 0.1267 & 0.2408 & 0.8158 \end{bmatrix}.$$

Simulation studies

A population subject to mass selection was simulated for 10 non-overlapping generations, subsequent to a base population of unrelated and non-inbred animals (generation 0). Each generation, 50 males and 200 females were

Table 1: Contributions to the mixed model equations in the case of fuzzy classification of genetic groups

Position	Contribution	
(i, i)	x	
(i, m_j), (m_j , i)	$-0.5 p_{ij} x$	$j = 1, \dots, s_i$
(i, f_k), (f_k , i)	$-0.5 p_{ik} x$	$k = 1, \dots, d_k$
(m_j , $m_{j'}$)	$0.25 p_{ij} p_{i'j'} x$	$j = 1, \dots, s_i$ $j' = 1, \dots, s_i$
(f_k , $f_{k'}$)	$0.25 p_{ik} p_{ik'} x$	$k = 1, \dots, d_k$ $k' = 1, \dots, d_k$
(m_j , f_k), (f_k , m_j)	$0.25 p_{ik} p_{ik'} x$	$j = 1, \dots, s_i$ $k = 1, \dots, d_k$

Table 2: Contributions to A* for animals A, B, C and D in the pedigree example

A	B	C	D
1.0 to A,A	1.0 to B,B	1.0 to C,C	2.0 to D,D
-0.5 to A,g ₁	-0.5 to B,g ₁	-0.15 to C,g ₂	-1.0 to D,A
-0.5 to g ₁ ,A	-0.5 to g ₁ ,B	-0.15 to g ₂ ,B	-1.0 to A,D
-0.5 to A,g ₂	-0.3 to B,g ₂	-0.35 to B,g ₃	-1.0 to D,B
-0.5 to g ₂ ,A	-0.3 to g ₂ ,B	-0.35 to g ₃ ,B	-1.0 to B,D
0.25 to g ₁ ,g ₁	-0.2 to B,g ₃	-0.5 to B,g ₃	0.5 to A,A
0.25 to g ₁ ,g ₂	-0.2 to g ₃ ,B	-0.5 to g ₃ ,B	0.5 to A,B
0.25 to g ₂ ,g ₁	0.25 to g ₁ ,g ₁	0.0225 to g ₂ ,g ₂	0.5 to B,B
0.25 to g ₂ ,g ₂	0.15 to g ₁ ,g ₂	0.0525 to g ₂ ,g ₃	0.5 to B,B
	0.15 to g ₂ ,g ₁	0.0525 to g ₃ ,g ₂	
	0.1 to g ₁ ,g ₃	0.075 to g ₂ ,g ₃	
	0.1 to g ₃ ,g ₁	0.075 to g ₃ ,g ₂	
	0.09 to g ₂ ,g ₂	0.1225 to g ₃ ,g ₃	
	0.06 to g ₂ ,g ₃	0.175 to g ₃ ,g ₃	
	0.06 to g ₃ ,g ₂	0.175 to g ₃ ,g ₃	
	0.04 to g ₃ ,g ₃	0.25 to g ₃ ,g ₃	

randomly mated. Each mating produced two offspring, one of each gender.

For each animal a phenotypic record was simulated as the sum of an overall mean, the animal's breeding value and a random residual. An animal's breeding value was generated as the sum of the parent average and a Mendelian sampling deviation that considered inbreeding of the sire and dam. Genetic and residual variances were both 10, yielding a heritability of 0.50.

For each replicate, a "real life", incomplete data set for genetic evaluation was created by randomly deleting data. When data were deleted, both phenotypic records and relationship were deleted. The probability of deletion decreased linearly with increasing generation number, and ranged between 0.30 (generation 10) and 0.70 (generation 0).

The model for genetic evaluation included an overall mean, random animal effect and genetic groups for phantom parents. The simulated variance components were used to predict breeding values. Solutions to the mixed model equations were obtained using the preconditioned gradient algorithm, which was assumed to be converged when the relative average difference between the right and left hand sides was smaller than 10⁻¹⁰.

Fuzzy classification was compared with distinct classification of genetic groups in two situations: 1) to handle the uncertainty about which group a phantom parent should be assigned to, and 2) to model the average genetic level of unknown parents with a small number of parameters.

Fuzzy classification to handle uncertainty

For the genetic evaluation and forming of genetic groups, uncertainty about the generation of birth was simulated for 25% of the animals with at least one unknown parent. The uncertainty was such that unknown parents could belong to two possible generations: the true generation of birth and the generation prior to that, each with equal probability (0.50). Phantom parents were grouped based on gender and generation number, which resulted in 20 different genetic groups. In the distinct classification, for animals with simulated uncertainty, phantom parents were randomly assigned to just one genetic group, either for the true generation of birth or the generation prior to that, with equal probability. In the fuzzy classification, these phantom parents were assigned to two genetic groups, for both possible generations, with equal proportions (0.50).

The simulation was repeated 50 times. For each replicate the empirical mean and variance of prediction errors (true minus predicted breeding value) were computed within generation for animals with simulated uncertainty. In addition, the rankings on predicted breeding values for distinct and fuzzy classification were compared with the rankings on true breeding values. Animals were grouped in deciles based on the predicted and true breeding value, and the percentage of animals classified in the correct decile was determined for animals with and without simulated uncertainty and for animals with both parents known.

Fuzzy classification for parsimonious modelling

Two strategies for assigning phantom parents to genetic groups were compared: distinct and fuzzy classification. In the distinct classification, phantom parents were grouped on the basis of gender and generation number, which resulted in 20 different genetic groups. The average number of animals per genetic group was 107 and group size ranged between 59 and 146. In the fuzzy classification there were four groups: two groups for the parent's gender (male, female) and two groups to describe the average genetic level of parents of generation 1 animals and one for parents of generation 10 animals. Phantom parents of animals in intermediate generations were assigned to both groups, with proportions depending on generation number. For example, a phantom sire of an animal born in generation two was assigned with 50% and 0% to the male and female parent genetic group and with 45% to the generation 1 genetic group and with 5% to the generation 10 genetic group, for a phantom dam of an animal born in generation three the proportions were 0%, 50, 40% and 10%, respectively, etc. This way, the sum of proportions was always equal to 1. The resulting Q matrix is not of full rank, meaning that only 3 degrees of freedom are used to model the genetic groups.

The simulation was repeated 50 times. For each replicate, the empirical mean and variance of prediction errors (true minus predicted breeding value) were computed within generations, separately for animals with 0, 1 or 2 unknown parents. The across-replicate standard deviation of estimates for each genetic group (20 and 3 in case of distinct and fuzzy classification, respectively) was also inspected as an indication of the SE of genetic group solutions. In addition, the across-replicate correlation between genetic group estimates was computed as an indication of sampling correlation of genetic group solutions.

Results

Fuzzy classification to handle uncertainty

The empirical prediction error variance was 1.6% lower ($P < 0.05$) for animals with simulated uncertainty when genetic groups were fuzzy-classified. For animals with unknown parent groups and for which the generation of birth was accurately known there was no difference in empirical prediction error variance between both grouping strategies. For animals with simulated uncertainty, the percentage of correctly ranked animals was higher (0.7 percent point; Table 3). More importantly, accounting for the uncertainty reduced the variation across replicates by more than 10% (Table 3).

The magnitude of the advantage of fuzzy classification is specific for this simulation design and will vary from case to case, whether it concerns other simulation designs or practical applications. The gains presented here represent a best case scenario, since some aspects of the pattern of uncertainty could be considered in the membership functions. In practical applications this will not be the case and there will be some noise even with fuzzy-classified groups.

Fuzzy classification for parsimonious modelling

The empirical prediction error variance was significantly ($P < 0.05$) lower when genetic groups were fuzzy-classified for animals with one unknown parent (Figure 2). The improvement ranged between 0.4% and 1.7%, depending on generation number. The SE of genetic group solutions was approximately 1.8 times lower for fuzzy-classified

groups compared to distinct genetic groups. In case of fuzzy classification, much fewer groups (about six times less) were needed to describe the linear pattern of genetic trend of unknown parents, resulting in a higher precision of the genetic group solutions and explaining the lower prediction error variance of breeding values for this alternative.

The correlation between solutions for genetic groups for phantom sires and dams (for the same generation) was moderately negative (~ -0.3) for the distinct classification of genetic groups, but slightly positive (~ 0.1) when groups were fuzzy-classified (results not shown). This result may explain why the prediction error variance differed between alternatives for animals with one phantom parent, but not for animals with two phantom parents. For an animal with two unknown parents, overestimation of the solution for the genetic group effect for one parent can be compensated by an underestimation of the solution for the genetic group effect for the other parent. When only one parent is unknown such compensation does not occur, and a more precise estimation of the genetic group effects, as in the fuzzy-classified alternative, is favourable.

Discussion

The rules for building the numerator relationship matrix in the case of fuzzy-classified groups are similar to those for building an average numerator relationship for cases where parenthood is uncertain, but limited to a small number of possible parents (see rules in [10,11], and [12]). The conceptual difference between both procedures is that for the average relationship potential parents are known, whereas in the case of (fuzzy-classified) genetic groups parents are unknown.

Incorporation of genetic groups in the genetic evaluation can have a substantial effect on the estimated genetic trend and selection of which parents to breed the next generation of animals (e.g. [13]). Inclusion of genetic groups may sometimes lead to incorrect ranking of animals and suboptimal selection decisions [14]. Therefore, for practical applications it is important to evaluate the consequence of different grouping approaches (distinct, fuzzy) and definition of genetic groups (on the basis of

Table 3: Percentage of animals correctly classified in deciles created on the basis of ranking on true breeding values.

Group of animals	Distinct classification		Fuzzy classification	
	Mean	SD	Mean	SD
Both parents known	39.7	1.4	39.6	1.4
Unknown parent(s) - simulated uncertainty	41.4	3.3	42.1 ^a	2.9 ^a
Unknown parent(s) - no uncertainty	41.1	3.3	41.2	3.1

^a Significantly different from distinct classification ($P < 0.05$)

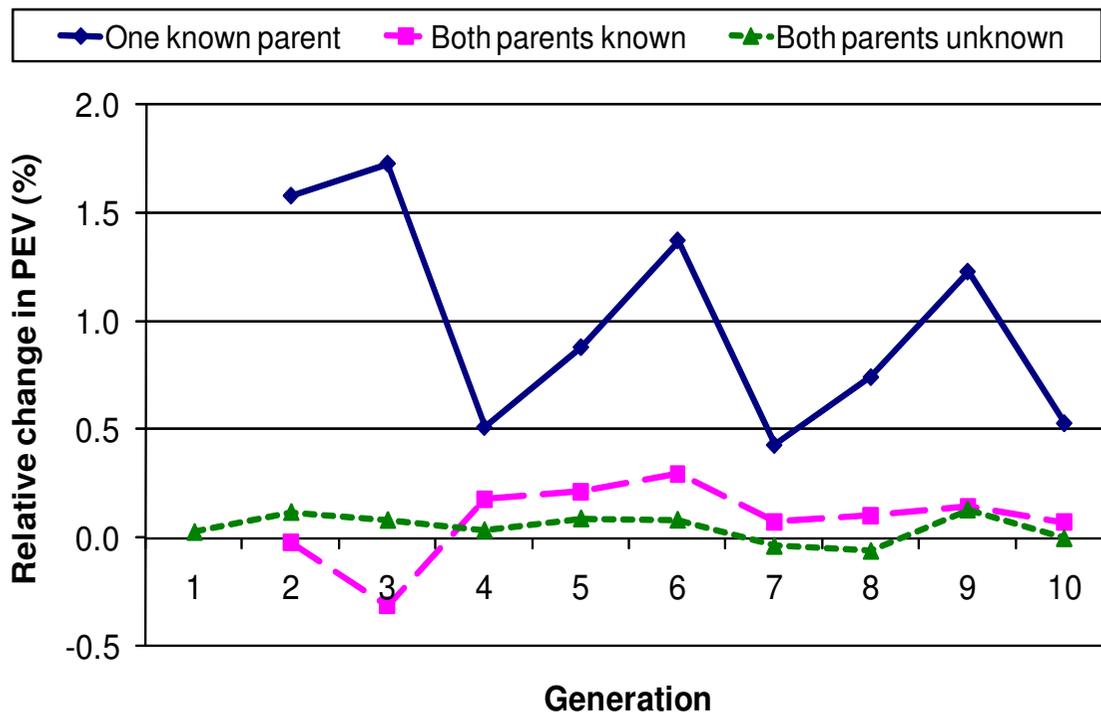


Figure 2
Relative difference (distinct vs. fuzzy classification; %) in prediction error variance for animals having 0, 1 or 2 phantom parent(s), by generation number.

birth year, selection path, origin, etc., or combinations thereof) before deciding which one to adopt.

Membership functions should follow the pattern of uncertainty and accurately describe the average genetic level of phantom parents as much as possible. However, consideration should also be given to the precision and the estimability of genetic group effects when determining the proportions with which phantom parents are assigned to genetic groups. For example, very small values are added to the diagonal elements of the mixed model equations for a genetic group to which phantom parents are assigned with small proportions, resulting in a high standard error for that genetic group. Also, to avoid confounding among genetic groups it is necessary to have several combinations of different genetic groups and for the same combination of genetic groups to use several different sets of membership proportions. Therefore, in the evaluation of fuzzy-classified genetic groups it is important to carefully examine the precision of genetic group estimates and possible confounding of genetic groups.

A linear trend in the average genetic level of phantom parents was modelled with a low number of parameters

(groups) by fuzzy-classification in the simulation. More complex modelling of the average genetic level of phantom parents is possible by means of the fuzzy-classification approach. For example, other factors, like origin and selection path, could be incorporated in a similar way.

Competing interests

The author declares that they have no competing interests.

Authors' contributions

WFF conceived and designed the study, programmed and carried out the computer simulations, analyzed and interpreted the results, and wrote the manuscript.

Acknowledgements

The valuable feedback of two anonymous reviewers is gratefully acknowledged.

References

1. Pollak EJ, Quaas RL: **Definition of group effects in sire evaluation models.** *J Dairy Sci* 1980, **66**:1503-1509.
2. Thompson R: **Sire evaluation.** *Biometrics* 1979, **35**:339-353.
3. Westell RA, Quaas RL, VanVleck LD: **Genetic groups in an animal model.** *J Dairy Sci* 1988, **71**:1310-1318.
4. Banos G, Schaeffer LR, Burnside EB: **Genetic relationships and linear model comparisons between United States and Cana-**

- dian Ayrshire and Jersey populations. *J Dairy Sci* 1991, **74**:1060-1068.
5. Robinson GK: **Group effects and computing strategies for models for estimating breeding values.** *J Dairy Sci* 1986, **69**:3106-3111.
 6. Quaas RL: **Additive genetic model with groups and relationships.** *J Dairy Sci* 1988, **71**:1338-1345.
 7. Strandberg E, Grandinson K: **Adjusting for seasonal effects in an animal model using fuzzy classification.** *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production: 11-16 January 1998; Armidale 1998*, **25**:633-636.
 8. Mrode RA: *Linear models for the prediction of animal breeding values* Wallingford, Oxon, UK: CAB International; 2005.
 9. Henderson CR: **A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values.** *Biometrics* 1976, **32**:69-83.
 10. Famula TR: **Simple and rapid inversion of additive relationship matrices incorporating parental uncertainty.** *J Anim Sci* 1992, **70**:1045-1048.
 11. Henderson CR: **Use of an average numerator relationship matrix for multiple-sire joining.** *J Anim Sci* 1988, **66**:1614-1621.
 12. Perez-Enciso M, Fernando RL: **Genetic evaluation with uncertain parentage: a comparison of methods.** *Theor Appl Genet* 1992, **84**:173-179.
 13. Theron HE, Kanfer FHJ, Rautenbach L: **The effect of phantom parent groups on genetic trend estimation.** *S Afr J Anim Sci* 2002, **32**:130-135.
 14. Phocas F, Laloë D: **Should genetic groups be fitted in BLUP evaluation? Practical answers for the French AI beef sire evaluation.** *Genet Sel Evol* 2004, **36**:325-345.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

