Modern Genetic Evaluation Procedures
Why BLUP?

Hans-Ulrich Graser

1 Introduction

The developments of modem genetic evaluation procedures have been mainly driven by scientists working with the dairy populations in the United States and Europe. The heavy use of artificial insemination allowed them to link data right across the country in one large data bank. Accurate evaluation of AI sires, which quite often have more than 100 000 daughters, was economically very important and sire evaluation become an art which cumulated in research leading to the development of Best Linear Unbiased Prediction Methods (BLUP), now widely used in many livestock species.

The Pig industry for many years had a different approach. This was due to some extend by the fact that AI was not readily adopted, and the performance testing system was more systematic and controlled, eg. central test stations. Pig breeders also wanted to improve a number of traits simultaneously and therefore used Selection Index combining the information of more than one trait through economic weights into one value on which to base selection decisions.

It was not for some years till it was generally understood, although well described in a paper in 1974, that the BLUP technology as developed by C.R. Henderson (Henderson, 1972) was in effect an unconditional selection index. The condition previously imposed was that the mean performance of all the recorded groups was known without error. BLUP removes this restrictions and accommodates the uncertainty about the mean (non genetic) performance of a group of animals.

Another factor which greatly assisted the development and use of BLUP models for the genetic evaluation of pig performance data was the ever increasing power of today's computers. While simple indices could be calculated by hand the selection index calculated with a multitrait animal model requires at least a standard PC.

2 What is so special about BLUP?

Consider the information about the soccer results of two teams presented in table 1.

Table 1. Soccer results of two teams

	P	W	D	L
Team 1	36	20	10	6
Team 2	36	10	12	14

While team 1 has won 20 games and lost only 6, team 2 lost 14 and won only 10 games. It seems obvious which is the better team! However we can not precisely answer this question without the knowledge in which division the two teams played. If we learn that team 1 played in the second division and team 2 in the first division, we will probably think twice before declaring team 1 to be the better one.

Similar complications apply to livestock breeding. Comparing animals to make genetic selection decisions requires knowledge about the genetic competition (division) of an animal and the environment (group) in which it has been compared in. Generally we try to exclude the environment by only comparing animals which have been treated alike, that means animals are of the same sex and age and have been performance tested together in the same shed. The genetic competition will be accounted for by the knowledge of the pedigree of the animal and the pedigree of its contemporaries.

By developing an equation system which contains both environmental and genetic effects at the same time we can simultaneously solve for genetic and environmental effects, adjusting one for the influence of the other.

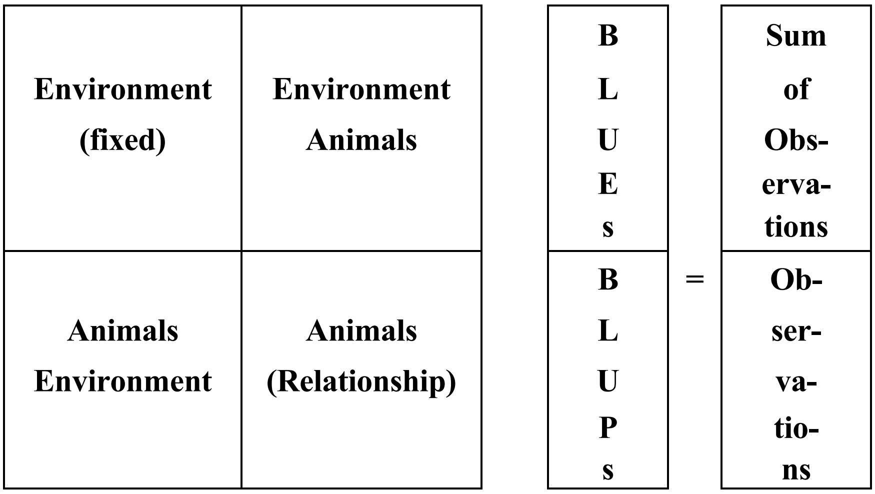

Figure 1. The BLUP Equation System

Therefore a good BLUP analyses results in

- correct treatment of unequal competition
- correct removal of selection bias from sequential selection
- correct incorporation of information of all relatives' information

Being able to link information across years and generations allows us to examine the effect of management decisions free of genetic differences and estimate the genetic trend free of environmental influences. All this is however only correct if we use the correct genetic parameters (heritabilities and correlations) in the analyses.

3 What does BLUP stand for?

BLUP is the abbreviation for the statistical term

B	Best
L	Linear
U	Unbiased
P	Predictor

This says simply that our Predictor (Estimated Breeding Value, EBV) has the smallest error (is best) of all unbiased predictors (expectation is equal to the mean result if we do it often enough) and is derived by a linear function of the data.

There might be other predictors which are biased (eg. underestimate the true value most of the time) but have a smaller error or are non linear. As animal breeders we prefer most of the time unbiasedness with larger errors, to bias with smaller errors, particularly if we can't identify the direction of the bias.

Figure 2. Example of bias and error (variance)

We can overcome large variance by "shooting" more often (have more information), bias can only be overcome if we know the direction.

4 Different BLUP Models

The use of BLUP technology for estimation of breeding values does not indicate how it is done. Models which describe the knowledge about the biology are required, e.g.

4.1 Sire Model

$$Y = M + AGE + SEX + sire + residual$$

where	Y	is an observation eg. Birthweight
	M	is the effect of the management group including year and season effects
	AGE	is the effect of the age of dam (older dams have heavier progeny)
	SEX	is the effect of the sex of the progeny (male are heavier than females)
	Sire	is the effect of the sire of the progeny, this is our main interest
	residual	the rest which we can not yet explain.

This model allows us to estimate half the breeding value of the sires provided we have good knowledge of the heritability of the trait that means we know the variance of the two random variables in our model sire effect and residual effect. These variances can be calculated if we know the phenotypic differences between animals and the heritabilities of the trait.

$$\sigma_s^2 = \tfrac{1}{4} h^2 \sigma_p^2, \qquad \sigma_e^2 = \sigma_p^2 - \sigma_s^2$$

Sire variance is a quarter of the additive genetic variance $\left(\sigma_a^2 = h^2 \sigma_p^2\right)$. Similar sire models were first used more than twenty years ago for the evaluation of AI dairy sires in the North Eastern States of the USA by the group around Prof Henderson which developed the first BLUP procedures.

An assumption made in this model is, that sires are randomly mated to the dams. If a group of sires is allocated heavier dams than another group of sires (eg. big heifers versus small heifers) we can not account for this selection bias unless we have it recorded. The best way to record it would be to identify dams and have their birth weight included in the analysis. This process leads via a couple of interim steps to the **Animal Model**.

The Animal Model is today becoming the standard model for genetic evaluation of all sorts of livestock: dairy and beef cattle, sheep, pigs, goats and poultry even dogs, racing pigeons and horses (dressage, show jumping and racing)

4.2 The Animal Model

$$Y = M + AGE + SEX + animal + residual$$

where		
where	animal	is the progeny's breeding value
	residual	is a different rest as in the sire model as we have explained more of the differences between the records

$$\left(\sigma_a^2 = h^2\ \sigma_p^2\right) \qquad \sigma_e^2 = \sigma_p^2 - \sigma_a^2$$

A short version of this animal model is

$$y = Xb + Za + e$$

where a = breeding value of animals
b = all fixed (other effects)
X, Z are called design matrices linking observations to effects

To estimate (predict) the breeding values we require knowledge about the distribution of the breeding values that means the variance and covariances of these breeding value. The variances and covariances (relationships) of the breeding values is described by the relationship matrix A multiplied with the genetic variance σ_a^2.

$$\mathrm{Var}(a) = A\,\sigma_a^2$$

4.3 The Relationship Matrix A

The following pedigree structure is used to demonstrate the information accumulated in a relationship matrix.

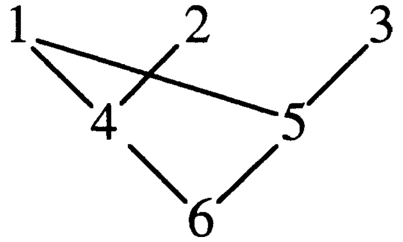

Table 2: Relationship Matrix between the six animals in the above pedigree showing the proportion of genes in common.

		Animal					
		1	**2**	**3**	**4**	**5**	**6**
A	1	1.0	0	0	0.5	0.5	0.5
n	2	0	1.0	0	0.5	0	0.25
i	3	0	0	1.0	0	0.5	0.25
m	4	0.5	0.5	0	1.0	0.25	0.625
a	5	0.5	0	0.5	0.25	1.0	0.625
l	6	0.5	0.25	0.25	0.625	0.625	1.125

The base animals (1,2,3) are unrelated (zero on the off diagonals) and not inbreed; their pedigree is unknown, therefore 1.0 on the diagonal. Animal 1 is a parent of both 4 and 5 and has passed half its genes to the progeny (0.5 on off diagonal). Animal 4 and 5 are half sibs and have a quarter of the genes from animal 1 in common. Animal 6 the product of the mating of the related animals 4 and 5 is inbred and the 1.125 on its diagonal reflects the 12.5% inbreeding. Animal 6 is of course related to its parents and grandparents.

This example is relatively trivial if you compare it to a population containing a million animals and many more generations of pedigrees. For example animals having a Great-Grand sire in common are related and this information would be reflected in a small off diagonal element between these animals. There would then be millions of off diagonal elements in such a matrix.

To predict the EBVs with the mixed model equations the relationship matrix A is not needed but its inverse A^{-1}. The late Prof C.R. Henderson who was the main "inventor" of BLUP also discovered that the structure of A^{-1} is considerably simpler than the structure of A (Henderson, 1975). The only information required to build A^{-1} is the inbreeding coefficient of the animals and the numbers of their parents. We can than follow simple rule to set up A^{-1} for use in the mixed model equations.

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + A^{-1} * G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

On the left hand sight of this equation system are the information about management groups, which animals are in which management group and how these animals are related plus the genetic correlations and heritabilities of the traits. In the middle are our unknown effects and breeding values which we want to estimate and predict, and on the right hand side are our records (measurements) summed over fixed effects (top part) and as single information for the animals. The only thing required now is some smart software to solve this equation system.

Remember how we solved three equations with three unknown at school. This is the same except there might be some tens of millions of equations, if we look at multitrait dairy evaluations. Of course our pig problems within a herd are smaller and the power of today's personal computers is sufficient to do the calculations within a reasonable time.

5 Summary

BLUP with an Animal Model is a modem genetic evaluation system which is relatively easy to understand, not necessarily easy to compute. By simultaneously solving for the environmental effects and the breeding values BLUP accounts for unequal competition, non random mating and sequential selection. However to allow BLUP technology to work best an accurate recording system has to be in place which keeps good pedigree records and stores information which is used for culling decisions.

References

Henderson C.R. 1972, "Sire evaluations and genetic trends", *Proc. of the Animal Breeding and Genetics Symposium in Honor of Dr. J.L. Lush.* Champaign , Illinois.

Henderson C.R. 1975, "Rapid method for computing the inverse of a relationship matrix", *J. Dairy Sci.*, 58, 1727-1730.

NOTES