

New Zealand Society of Animal Production online archive

This paper is from the New Zealand Society for Animal Production online archive. NZSAP holds a regular annual conference in June or July each year for the presentation of technical and applied topics in animal production. NZSAP plays an important role as a forum fostering research in all areas of animal production including production systems, nutrition, meat science, animal welfare, wool science, animal breeding and genetics.

An invitation is extended to all those involved in the field of animal production to apply for membership of the New Zealand Society of Animal Production at our website www.nzsap.org.nz

[View All Proceedings](#)

[Next Conference](#)

[Join NZSAP](#)

The New Zealand Society of Animal Production in publishing the conference proceedings is engaged in disseminating information, not rendering professional advice or services. The views expressed herein do not necessarily represent the views of the New Zealand Society of Animal Production and the New Zealand Society of Animal Production expressly disclaims any form of liability with respect to anything done or omitted to be done in reliance upon the contents of these proceedings.

This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](http://creativecommons.org/licenses/by-nc-nd/4.0/).



You are free to:

Share— copy and redistribute the material in any medium or format

Under the following terms:

Attribution — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

NonCommercial — You may not use the material for [commercial purposes](#).

NoDerivatives — If you [remix, transform, or build upon](#) the material, you may not distribute the modified material.

<http://creativecommons.org.nz/licences/licences-explained/>

Best linear unbiased prediction for across-flock/year breeding values

D. J. GARRICK

Animal Science Department, Massey University, Palmerston North, New Zealand.

ABSTRACT

In New Zealand, breeding values have been predicted using the method of selection index (Sheeplan, Beeplan, Animalplan) or variants such as a modified contemporary comparison (Dairy Cattle). Over the last decade, research systems for evaluating beef cattle, dairy cattle, sheep and pigs in New Zealand have increasingly made use of Henderson's mixed model equations. In appropriate circumstances, these equations give rise to evaluations which are best, linear and unbiased predictions (BLUP). This review compares and contrasts theoretical and applied aspects of evaluation using selection index and BLUP. An example is used to demonstrate the rationale of BLUP for evaluating animals with records in different contemporary groups such as across-flocks or years.

Keywords BLUP, selection index, sire reference analysis.

INTRODUCTION

For several decades, farmers in New Zealand have had access to sires which have been objectively selected for improved performance. Objective ranking of potential candidates to parent the next generation has occurred through collection and statistical analysis of performance data. The statistical method most widely used in New Zealand for ranking animals is known as Best Linear Prediction (BLP) or Selection Index. This technique is often inappropriately applied for comparison of animals with performance records obtained in different flocks (or years) for which the genetic and environmental means are not known. A technique known as Best Linear Unbiased Prediction (BLUP) is often appropriate for comparing animals with performance records obtained in more than one flock or herd environment. Animal breeding researchers in New Zealand have made use of BLUP techniques for at least 18 years (Anderson, 1974; Blair, 1981) and since 1985 it has been available for commercial use to obtain across-flock ranking of sheep (in group breeding and sire reference schemes). The objective of this paper is to compare and contrast the techniques of BLP and BLUP to provide some insight as to why BLUP is a desirable approach for comparing animals across flocks or herds.

GENETIC EVALUATION

The development of a statistical procedure for evaluating animals starts with the definition of a model. The model includes an equation to summarise our knowledge of biology of the trait(s) of interest. In many cases, the equation is assumed linear (perhaps after transformation). For example

$$y_{ijk} = \mu_i + g_j + e_{ijk} \quad [1]$$

where

y_{ijk} denotes the jk^{th} observation from an animal in the i^{th} subclass (e.g. flock-year), μ_i represents the mean level of performance for subclass i , g_j denotes the additive genetic component relating to the performance record, and e_{ijk} comprises the residual effect, including environmental and non-additive genetic effects.

For notational convenience, model equation [1] can be specified for every observation and accumulated in matrices and vectors (represented by bold letters) in the following form.

$$y = X\beta + Zg + e \quad [2]$$

Thus, y is a vector of all observations, β is a vector containing all required means, X is a known model matrix relating means in β to observations in y , g is a vector containing genetic effects with the known model matrix Z relating genetic effects to observations and e is a vector of residual effects, one effect corresponding to each observation in y . The model equations in [1] and [2] can be used to define a sire model (where g_j denotes the genetic contribution of the sire of the animal with observation y_{ijk}) or an animal model (where there is one g_j corresponding to every observation).

The model is not completely specified until distributional properties of the effects in the model equation are defined. The usual definitions for a single trait model involve specification of β as a vector of fixed effects (such as flock means, age adjustments), g as a vector of random effects with variance-covariance matrix, $G = A\sigma_g^2$ (A denotes the additive relationship matrix) and e as a vector of random effects with variance-covariance matrix, $R = I\sigma_e^2$ (I is an identity matrix, implying each residual effect has the same variance and residual effects are mutually uncorrelated). Finally, g and e are assumed to have zero covariance, equivalent to assuming no genotype-environment interaction. These models which include effects which are fixed (such as β) and effects which are random (other than the residual) are known as mixed models.

The goal of genetic evaluation is to estimate breeding values (denoted by \hat{g}) from a mathematical function of the observations that will provide a predictor of g , which is in some sense optimal for selection. In the prediction sense, the usual definition of optimal (Best) is the minimum prediction error variance. The prediction error variance defines the average magnitude of the square of the difference between an estimate of g_j and its true value.

Best Linear Prediction

The first widely-used approach for predicting genetic effects in [1] and [2] was based on obtaining the Best (minimum prediction error variance) Predictor, with predictors limited to *Linear* functions of observations (hence BLP). It may seem limiting to restrict ourselves to linear predictors, however it can be shown, for

models with g and e following a normal distribution, the Best Predictor is linear (Henderson, 1973). The method is also widely known as selection index and is taught in many undergraduate animal breeding courses. Derivation of BLP requires that first and second moments of the distributions (i.e. the means in β and variances σ_g^2 and σ_e^2) are known without error. In this case, [2] can be rearranged

$$y - X\beta = Zg + e \quad [3]$$

and the predictor is

$$\hat{g} = (Z'R^{-1}Z + G^{-1})^{-1}Z'R^{-1}(y - X\beta) \quad [4]$$

which in the single-trait case simplifies to

$$\hat{g} = (Z'Z + \alpha A^{-1})^{-1}Z'(y - X\beta) \quad [5]$$

with

$$\alpha = \sigma_e^2 / \sigma_g^2$$

Further simplifications are possible by ignoring certain relationships to simplify the structure of A^{-1} . Operationally, an evaluation involves adjusting the data for fixed effects, then multiplying the adjusted data by predetermined weighting factors. The problem arises that in practice, neither the fixed effects (to preadjust the data) or variance parameters (to obtain α) are known without error. Research has demonstrated that rankings of animals are not changed markedly when an estimate of α , not equal to the true value, is used. However, expected economic response may differ markedly from actual response. The use of estimates of fixed effects to adjust data will bias estimated breeding values upwards (if adjustments are too small) or downwards (if too large an adjustment is used). This bias will not be serious when estimates are close to true parameter values (e.g. adjustment for within contemporary group effect of age of lamb on weaning weight). For flock-year or herd-year means, large datasets are not usually available and adjustments will have errors which cause bias and reduce selection efficiency. A further problem with BLP is that it is sometimes desirable to use all information on available

relatives which leads to different weighting factors from [5] for every animal.

Mixed Model Equations for BLUP

In the late 1940's, Dr Henderson of Cornell University developed a set of equations for predicting random effects without assuming parameter values of fixed effects to be known. The equations were similar to [4] although the properties of the predictors were unknown. In 1955-56, while Henderson was on leave with the New Zealand Dairy Board, a young collaborator (S.R. Searle) discovered that solutions to the fixed effects from Henderson's equations had some optimal properties (Henderson *et al.*, 1959). This led to the discovery that predictions of random effects, from what are now known as the Mixed Model Equations (MME), are Best, Linear and Unbiased (BLUP) (Henderson, 1963). Assuming variance parameters are known, a form of the MME for single trait analysis corresponding to [2] are

$$\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \alpha\mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta}^0 \\ \hat{\mathbf{g}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix} \quad [6]$$

Examination of an appropriate row of [6] leads to the following predictors.

(i) Individual j with performance record y_{ijk} and unknown parents.

$$\begin{pmatrix} \text{adjustments} \\ \text{for fixed} \\ \text{effects of} \\ \text{subclass } i \end{pmatrix} + \hat{g}_j + \alpha\hat{g}_j = y_{ijk} \quad [7]$$

rearranging to

$$\hat{g}_j = h^2 \quad [8]$$

for $h^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \sigma_e^2)}$, known as heritability.

(ii) Individual j with performance record y_{ijk} and both parents included in $\hat{\mathbf{g}}$.

$$\begin{pmatrix} \text{adjustments} \\ \text{for fixed} \\ \text{effects of} \\ \text{subclass } i \end{pmatrix} + \hat{g}_j - \hat{g}_{\text{sire}} + \hat{g}_{\text{dam}} + 2\alpha\hat{g}_j = y_{ijk} \quad [9]$$

and after algebra

$$\hat{g}_j = \frac{h^2}{2-h^2} \frac{2(1-h^2)}{2-h^2} \frac{(\hat{g}_{\text{sire}} + \hat{g}_{\text{dam}})}{2} \quad [10]$$

That is, the predicted breeding value of an animal is a weighted function of the average breeding values of the parents and the animal's own adjusted deviation. The weights sum to one and the relative emphasis on parental or individual information is a function of the heritability of the trait.

The equations in [8] and [10] are exactly the same as BLP predictors except that the performance record is adjusted for fixed effects from the partition of [6] rather than preadjusted by known parameter values. This suggests BLUP evaluations result in adjustments calculated from the same dataset used for predicting genetic merit.

Examining the partition of [6] corresponding to fixed effects demonstrates the calculation of the adjustment factors.

$$\mathbf{X}'\mathbf{X}\hat{\beta}^0 + \mathbf{X}'\mathbf{Z}\hat{\mathbf{g}} = \mathbf{X}'\mathbf{y}$$

giving

$$\hat{\beta}^0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y} - \mathbf{Z}\hat{\mathbf{g}}) \quad [11]$$

equivalent to the least squares solutions for fixed effects except the righthand side is preadjusted for the estimated merit of the animals with the records. From [8] and [10] it can be seen that predictions of breeding values require solutions for fixed effects and [11] demonstrates that solutions for fixed effects require predictions of breeding values. Numerically, solutions to random and fixed effects must be obtained simultaneously, leading to greater computational effort than with [5]. Given that simultaneous solution is required, the inclusion of all known relationships makes little difference to computational effort, in contrast to BLP in [5] which could be readily simplified by ignoring most or all

relationships between animals being evaluated.

Equivalent Random Model for obtaining BLUP

Equations [6] and [11] provide little insight into the means with which BLUP provides predictions of breeding values that will provide greater response to selection relative to BLP. The following section develops an equivalent random model for across-flock/herd sire analysis, that gives rise to predictions of genetic merit which are identical to BLUP. It is best demonstrated with an example. Suppose we have performance observations from two flocks with two sires per flock, one (reference) sire having progeny in both flocks. Of particular interest is the comparison of sire 1 with sire 3.

	Sire Tag 1 Mean (no)	Sire Tag 2 Mean (no)	Sire Tag 3 Mean (no)
Environment			
Flock/Herd 1	$\bar{y}_{11} \cdot (n_{11})$	$\bar{y}_{12} \cdot (n_{12})$	
Flock/Herd 2		$\bar{y}_{22} \cdot (n_{22})$	$\bar{y}_{23} \cdot (n_{23})$

According to model equation [1], the progeny mean \bar{y}_{11} includes the effect of flock subclass 1. Similarly, progeny mean \bar{y}_{23} includes the fixed effect of flock subclass 2. Comparing these two progeny means provides little useful information unless the fixed flock effects are known or at least good estimates are available. Using BLP, flock effects would be estimated in some manner, with estimates represented by $\mu_1^0 = \mu_1 + \varepsilon_1$ and $\mu_2^0 = \mu_2 + \varepsilon_2$ respectively, denoting the error in estimating the effect of subclass *i*. Subtracting adjusted deviations, $(\bar{y}_{11} - \mu_1^0) - (\bar{y}_{23} - \mu_2^0)$, will estimate true genetic difference between sires *plus* bias from errors in estimating flock effects; i.e. $(g_1 - g_3) + (\varepsilon_2 - \varepsilon_1)$. In respect to equation [5], assuming the three sires are unrelated, the relevant matrices are as follows.

$$ZZ = \begin{pmatrix} n_{11} & 0 & 0 \\ 0 & n_{12}+n_{22} & 0 \\ 0 & 0 & n_{23} \end{pmatrix}, A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$Z'(y - X\beta)$ is estimated by

$$\begin{pmatrix} n_{11}(\bar{y}_{11} - \mu_1^0) \\ n_{12}(\bar{y}_{12} - \mu_1^0) + n_{22}(\bar{y}_{22} - \mu_2^0) \\ n_{23}(\bar{y}_{23} - \mu_2^0) \end{pmatrix} \quad [12]$$

The idea of BLUP is that appropriately chosen linear functions of observations provide information for evaluating random effects regardless of the value of unknown fixed effects. For example, using only performance records from flock 1, the difference between means,

$$\bar{y}_{11} - \bar{y}_{12} = (\mu_1 + g_1 + \bar{\varepsilon}_{11}) - (\mu_1 + g_2 + \bar{\varepsilon}_{12}),$$

has expected value $g_1 - g_2$. No further information is available on the comparison of sire 1 with sire 2 unless the flock effects can be removed to allow records from flock 2 to improve the evaluation of sire 2. Similarly, $\bar{y}_{22} - \bar{y}_{23}$ has expected value $g_2 - g_3$ and provides a comparison of sire 2 with sire 3 using only records from flock 2. Sire 2 can then provide a reference for comparing sire 1 with sire 3. That is, sire 1 can only be compared to sire 3 through the *connection* provided by sire 2. In general, the first step of a BLUP analysis is to discard data from flock-years which do not provide *connectedness* with other flock-years in the analysis. The idea of a sire reference scheme is to provide these links by reference sires although any sire linking flock-years will provide connectedness. In theory, BLP does not require connectedness because flock-year effects are assumed known such that data can be preadjusted to provide records as if all records had come from a single group of contemporaries.

Originally, four means were available for predicting the merit of the three sires. A BLUP analysis makes use of two differences between means to predict the random genetic effects. Some information has been

lost in transforming four means to two differences, but the data not used is informative for estimating flock effects only and not sire effects. As there were two flock effects in the analysis, two items of information were not useful for predicting random effects. Transforming the original data into differences between observations, such that the differences do not involve the fixed effects, allows a fully random model to be fitted, with identical predictions of breeding value as would be achieved by fitting mixed model equations as in [6]. The following matrices, used in [4] will give BLUP estimates of sire effects (N.B. Breeding value of a sire is twice the sire effect).

$$\mathbf{Z} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{R}^{-1} = \begin{pmatrix} \frac{\mathbf{n}_{11}\mathbf{n}_{12}}{\mathbf{n}_{11} + \mathbf{n}_{12}} & 0 \\ 0 & \frac{\mathbf{n}_{22}\mathbf{n}_{23}}{\mathbf{n}_{22} + \mathbf{n}_{23}} \end{pmatrix}, \alpha = \frac{(4 - h^2)}{h^2}$$

and using

$$\mathbf{y}^* = \begin{pmatrix} \bar{y}_{11} & \bar{y}_{12} \\ \bar{y}_{22} & \bar{y}_{23} \end{pmatrix}$$

in place of $(\mathbf{y} - \mathbf{X}\beta)$. This results in a matrix to be inverted from [4] which includes off-diagonal nonzero elements. Accordingly, there is a prediction error covariance between predicted breeding values for each of the three sires. This was not the case for the diagonal coefficient matrix from BLP which resulted in independent predictions of sire merit on the assumption that flock effects were known without error.

BLP vs BLUP

In theory, the techniques (i) BLP and (ii) BLUP are each optimal for their own situation, namely prediction using linear functions of observations for (i) fully random models or mixed models with known fixed effects and (ii) mixed models including unknown fixed effects, respectively. The use of BLUP when parameter values for fixed effects are known will result in less accurate evaluations than using BLP. If fixed effects are not

known, changing from currently used BLP techniques to BLUP will improve selection efficiency in two ways. First, the performance of all relatives can be used for predicting the value of any one animal. The improved accuracy of prediction from using all relatives may be minor for animals with their own performance record on traits with moderate heritability. Second, BLUP will provide estimates of genetic merit which are not biased by adjustment for contemporary group effects such as years or flock/herd environments. This will only be of advantage if candidates available for selection come from different contemporary groups, such as when selection is among animals born in different years or from different flock/herds.

Application of BLUP has lagged behind theory by at least 10-20 years, partly because of an imagined expense in the setting up and simultaneous solution of a large number of equations. In recent years, improved algorithms have become available which enable micro-computers to undertake BLUP evaluations which were considered impractical on mainframe computers 15 years ago. Mainframe computers are now being routinely used for national dairy evaluations using an animal model, whereby effects are included for every herd-year and animal with available data. Resulting equations in [6] may involve effects numbering tens of millions. A programme is currently being initiated to predict BLUP breeding values for dairy cattle in a single international analysis. Research to estimate variance components (and related heritabilities and correlations) from mixed models using the preferred likelihood methods (e.g. REML) require repeated BLUP solutions. Thus animal breeding research often involves development of a BLUP evaluation system as a prerequisite to estimation of variance components. Given the price and power of modern computers, the major cost of BLUP evaluation systems are in one-off software development.

A question which has been considered important in New Zealand is how much improvement in response to selection will be obtained by replacing current selection index systems with BLUP. Fully answering this question is probably more complex and just as costly as implementing BLUP. A comparison of dairy sire evaluation procedures on a theoretical (Dempfle and Haggar, 1983) and empirical basis (Haggar and Dempfle, 1983) demonstrated an improvement of 1-5%

in genetic gain from replacing a contemporary comparison with BLUP. The 5% improvement in response from BLUP occurred when selection of sires was across groups of herds with groups having different average genetic merit. The use of artificial insemination in the dairy industry results in progeny of each sire being distributed over more herds than would be the case for New Zealand sire reference schemes involving sheep, beef cattle, goats or deer. In a simulation study over 20 generations using multiple-trait within-herd selection of beef cattle, Robinson *et al* (1989) demonstrated BLUP resulted in almost 19% greater response in net worth. A study of swine selection strategies (Mabry and See; 1990) reported a 16% gain in response by using BLUP to select across-contemporary groups rather than retaining a fixed proportion from each group of pigs tested. A further 38% improvement in response was obtained by using artificial insemination to allow BLUP selection of boars across herds. A study of pig selection in Denmark (Blair; pers. comm.) determined a 20% improved response using BLUP compared to selection index. About one-quarter of the improvement resulted in increased accuracy from including all relatives in the BLUP evaluation, the remaining three-quarters resulted from removal of bias when selecting animals from different contemporary groups.

CONCLUSION

Strictly speaking, there is no choice among these techniques. If parameter values for fixed effects are not known, predictions should be based on mixed model equations. In practice, genetic evaluation of livestock species always involves fixed effects for contemporary groupings such as flocks/herds and years. In most species effects for age of the animal (often a covariate) or age of the dam may be important. Species with multiple progeny per parity are often affected by birth and/or rearing ranks. The crucial question to be asked in deciding between BLP and BLUP (ignoring possible

minor cost differences) is whether selection is to take place within or across contemporary groups. For example, a commercial sheep flock obtaining replacements from within-flock could carry out selection within- or across-years. If a predetermined age structure is to be used such that most selection pressure will occur in one age group such as two-tooth selection, then BLUP will likely provide little advantage over BLP other than providing useful information on retrospective genetic progress. However, if selection is to be practised across age-groups (years) and across flocks, then BLUP is the technique of choice.

REFERENCES

- Anderson, R. D. 1974. A study of biases in dairy sire evaluation. *M. Agr. Sc. thesis*, Massey University.
- Blair, H. T. 1981. Response to selection for open face and greasy fleece weight in Romney sheep. *Ph. D. thesis*, Massey University.
- Dempfle, L.; Haggar, Ch. 1983. Comparison of the efficiency of BLUP with other estimation procedures in dairy sire evaluation. I. Theoretical investigations. *Sonderdruck aus Zeitschrift fur Tierzuchtung und Zuchtungsbiologie 100*: 196-208.
- Haggar, Ch.; Dempfle, L. 1983. Comparison of the efficiency of BLUP with other estimation procedures in dairy sire evaluation. II. Empirical investigations. *Sonderdruck aus Zeitschrift fur Tierzuchtung und Zuchtungsbiologie 100*: 266-270.
- Henderson, C. R. 1973. Sire evaluation and genetic trends. In: *Proceedings of the Animal Breeding and Genetics Symposium held in Blacksburg, Virginia* 10-41. American Society of Animal Sciences, Champaign, Illinois.
- Henderson, C.R. 1963. Selection Index and expected genetic advance. *National Academy of Science - National Research Council* pp 982.
- Henderson, C. R.; Kempthorne, O.; Searle, S. R.; von Krosigk, C. M. 1959. The estimation of environmental and genetic trends from records subject to culling. *Biometrics 15*: 192.
- Mabry, J. W.; See, T. 1990. Selection with the animal model versus selection within contemporary groups for swine. *Journal of Dairy Science 73*: 2657-2665.
- Robinson, J. A. B.; Wilton, J. W.; Schaeffer, L. R. 1989. Accuracy of selection index and best linear unbiased prediction for within-herd selection with assortative mating of beef cattle. *Canadian Journal of Animal Science 69*: 315-322.