PIGBLUP Version 5.10

Ron Crump and Tony Henzell

Animal Genetics and Breeding Unit, a joint venture of NSW Agriculture and The University of New England.

Changes in PIGBLUP 5.10

Version 5.10 of PIGBLUP builds on the improvements to the user-interface of version 5.00 and adds new functionality. Particularly noteworthy is the inclusion of accuracy calculations. This is the first addition to PIGBLUP that has resulted from developments for the National Pig Improvement Program.

Version 5.10 contains three major additions to PIGBLUP's analytical capabilities and important extra options in PIGCHECK.

For PIGBLUP, the new analyses are:
- Automatic calculation of accuracies,
- Definition, Selection and Calculation of Multiple-Indexes plus the addition of a new economic model as the basis for a new index, and
- An advanced selection and mate allocation module.

In addition to these new capabilities, V5.10 allows:
- Analysis of ultra-sonic muscle depth measurements for a Live Muscle Depth (LMD) trait. For some time there has been a field for this trait included in record type two of the PIGBLUP data file format and this has now been activated.
- An extended batch description field for group formation when analysing IGF1. When the PIGBLUP data format was modified to include IGF1 records (record type 5), the batch code format (two characters) was modelled on that used in the early trials. PrimeGro routinely generate a longer code than this, requiring re-coding to reduce it to a two-character string. Now the old batch code slot can be used, or the full PrimeGro code can be passed through into a field on the end of record type 5.
- Viewing of various combinations of estimated breeding values (EBVs) and accuracies.
- Graphing of class statistics.
- Where multiple breeds are analysed, the ability to display genetic or environmental trends for all traits within a selected breed or all breeds within a selected trait.
- A new and more capable Migrate program to assist in upgrading PIGBLUP versions.

For PIGCHECK, the new options are:
- Detection of pedigree errors, automatic generation of suggested corrections, manual editing of corrections, and, if the user approves, application of the corrections to the data extract.
- Re-sorting of a data extract to ensure all records are correctly ordered for a PIGBLUP analysis.

Accuracies

1. Background

Estimated breeding values (EBVs) are a measure of the genetic merit of an animal within the population analysed. The more information available on an animal, the better its EBVs will be predicted. The accuracy is a measure of the quantity and quality of information contributing to each EBV. As the accuracy increases the probability that the estimated breeding value is different from the true breeding value decreases.

Please note: the accuracy of an EBV is taken account of in BLUP evaluation. You should not select, or pre-select, on accuracy.

The accuracy of an EBV for a trait is affected by a number of variables including:

- Heritability

  The heritability is the proportion of the phenotypic variation for a trait that is due to genetics. With this in mind, it is apparent that if the heritability of a trait was equal to one the record itself would precisely indicate the genetic merit of the animal. Similarly, if it were zero the record would provide no information about the genetic merit. Figure 1 shows the relationship between heritability and accuracy when only a single record on the animal itself (in a very large contemporary group) is available.
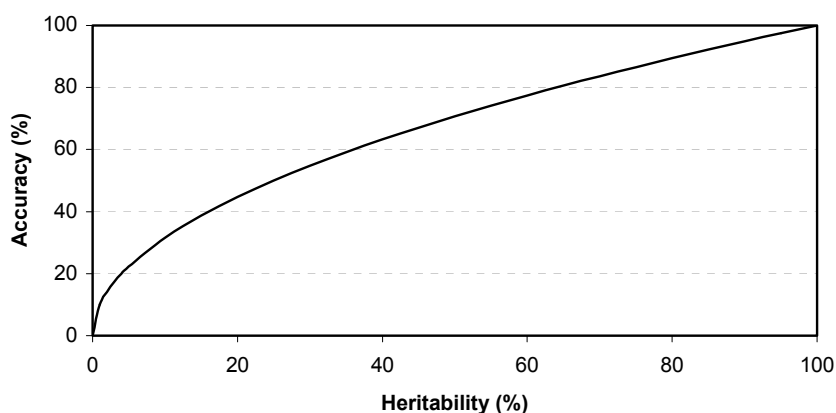

Figure 1 The relationship between accuracy and heritability with a single record on the individual.

  As heritability increases, the importance of a record on the individual increases and the importance of records on relatives decreases.

- Contemporary group size

  The more contemporaries are recorded together, the better the estimate of the management effect associated with them and hence the better the accuracy of the resulting EBVs.

  Figure 2 shows the impact of contemporary group size upon the accuracy of an EBV for different heritabilities. In this example the only information available

on an animal is its own record in a contemporary group of a certain size. A contemporary group containing only the animal itself provides no information for the evaluation. The accuracy rapidly asymptotes as the contemporary group size increases, that is to say that each additional piece of information increases the accuracy by a smaller amount.
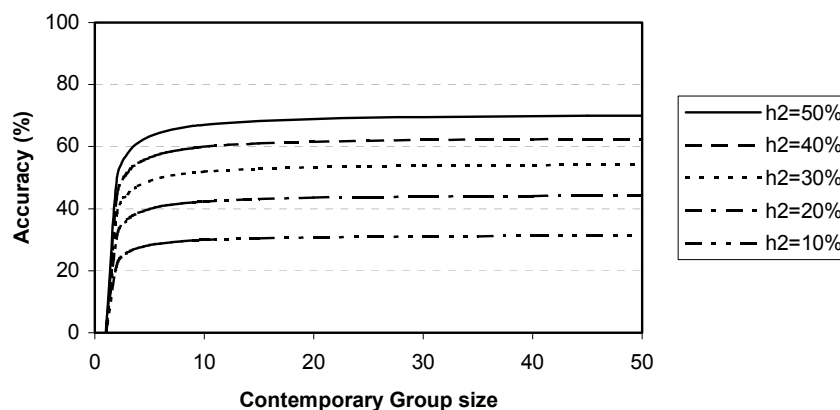

Figure 2 The effect of contemporary group size on accuracy at different levels of heritability. An individual only has a single record on the trait in a contemporary group of the given size.

- Information on relatives

    In an animal model BLUP procedure, such as in PIGBLUP, all available information on all known relatives is taken into account. Because more closely related animals are expected to have more genes in common and hence be more similar genetically, evaluations based on close relatives will be more accurate than those based on more distant relatives.

    The accumulation of information is not linear, ie information on the first offspring of a boar contributes more to the accuracy of his EBVs than information on the 25th offspring of this boar. Figure 3 shows the accuracy of a boar with one record himself plus records on n half-sib offspring. Contemporary group size was assumed to be large, so the accuracy for the boar when $n=0$ is the square root of the heritability of the trait.
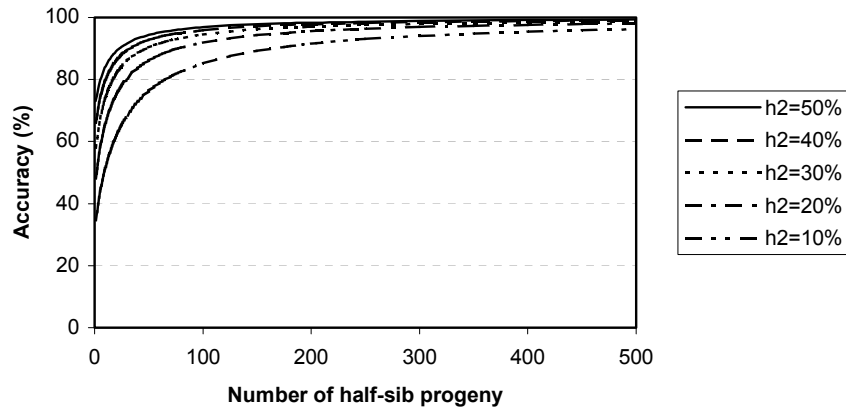
Figure 3 Accuracy of an EBV for a boar with his own record plus *n* recorded half-sib progeny.

- Information on other traits

    Correlated traits in the evaluation contribute to the accuracy for the trait being looked at. Illustration of the effects of correlated traits on the accuracy is less straightforward than for univariate sources of information, as there are multiple parameters interacting. Figure 4 shows the accuracy of the EBV for a trait when the EBV is estimated from the animal's own records for the trait of interest plus a second trait. The heritability and phenotypic standard deviation of both traits are 25% and 1 unit, respectively, and the phenotypic correlation was fixed at 0. As the magnitude of the genetic correlation increases and all other variables remain constant, the second trait provides more information. Note that the sign of the correlation does not affect the amount of information provided.
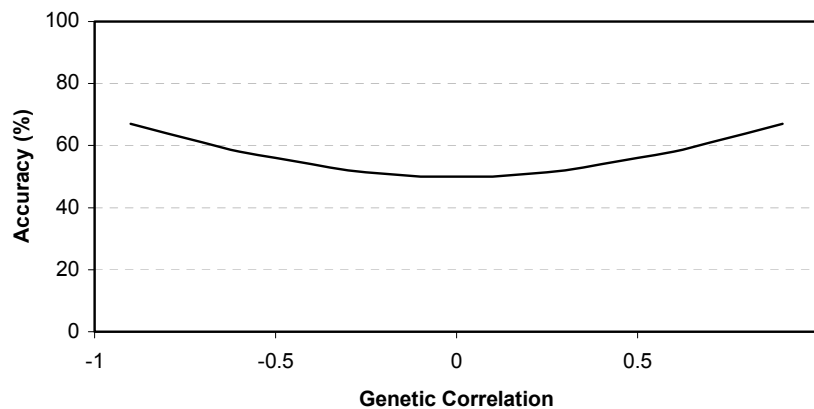

Figure 4 Accuracy of an EBV on a single trait, predicted from the animal's own records on the trait of interest plus a correlated trait. For more information refer to text.

- Information from repeated records

  For some traits repeated records are available from an individual. Within PIGBLUP, analysis of repeated records is limited to number born alive (NBA) and 21-day litter weight in all parities after the first (gilt parity records are treated as a separate trait). In Figure 5 the accuracy for NBA1 (first parity NBA), NBA2 (subsequent parity NBA) and NBAC (NBA1 + 1.5×NBA2) is presented, using PIGBLUP default genetic parameter settings. The sow in question has a single record on NBA1 then *n* subsequent parity records; no other information (eg on correlated traits or relatives) is available. When there is no record on NBA2, NBA1 is most accurate and the accuracy for NBAC reflects the NBA1 information. As records on NBA2 accumulate, the accuracy of NBA2 increases, and there is an increase in the accuracy of NBA1 via the high genetic correlation and the NBAC accuracy tends towards that of NBA2, due to the much larger amount of information on this trait.
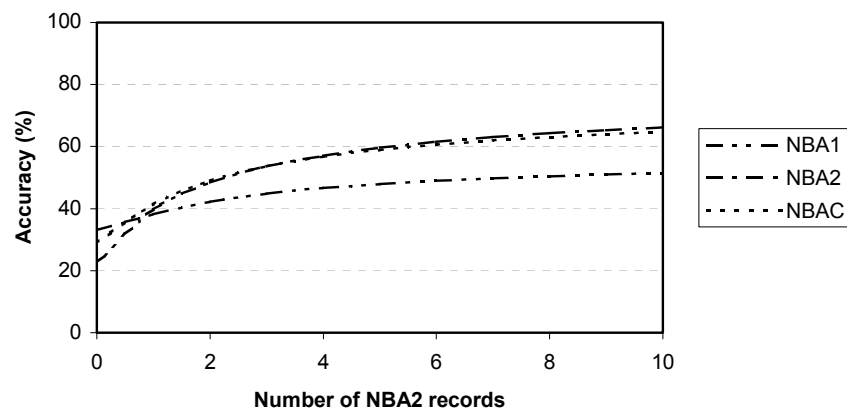

Figure 5 Accuracy of a sow's EBVs for NBA1, NBA2 and NBAC with increasing numbers of NBA2 records.

Each of the examples given only illustrates a very restricted case. In a real evaluation all of the information sources and the genetic parameters will be interacting with one another.

2. Implementation in PIGBLUP 5.10

Best linear unbiased prediction (BLUP) provides solutions to a set of equations. These equations can be represented as $\mathbf{y} = \mathbf{Cu}$, where $\mathbf{y}$ contains information on observations, $\mathbf{u}$ contains the effects that will be estimated and $\mathbf{C}$ contains all the necessary information to link observations with effects. As a very simple example if I paid $2 (my $\mathbf{y}$) for 5 ($\mathbf{C}$) apples, then by division apples cost 40c ($\mathbf{u}$) each on average. With the quantity of effects to be estimated in a typical realistic BLUP evaluation, $\mathbf{C}$ is a large matrix. The equivalent of dividing $\mathbf{y}$ by $\mathbf{C}$ when $\mathbf{C}$ is a matrix and $\mathbf{y}$ is a vector is to invert $\mathbf{C}$ and then multiply $\mathbf{y}$ by it. Unfortunately, inverting a large matrix is hard work for a computer, requiring a lot of calculations and much memory. As a result, BLUP runs are generally carried out using an iterative process which only requires $\mathbf{C}$ rather than its inverse (you can see this occurring on the progress screen in PIGBLUP, with a

convergence criterion decreasing as the solutions are approached). What does this have to do with accuracies? The accuracy of the solution for any of the effects in **u** is calculated from the diagonal element of the inverse of **C** that corresponds to that effect, so being unable to routinely invert **C** means that we also have to use an approximation of the accuracy.

The approximation used in PIGBLUP is based on Graser and Tier (1997) and was originally implemented by Horst Brandt while at AGBU and has been used in the NPIP for some time. The algorithms have been modified to work with PIGBLUP intermediate data files and to calculate approximate accuracies for traits and models as defined in PIGBLUP (eg with multivariate reproduction traits rather than a univariate repeatability model). This approximation does not take account of contemporary group size. PIGBLUP provides flexibility in assigning contemporary groups to ensure that there are reasonable numbers in each group and then, in accordance with Figure 2, these become irrelevant.

The approximation estimates the diagonal elements of the inverted coefficient matrix only. These diagonals relate to the variance of the predicted effect. The approximation does not calculate off-diagonals, ie covariances amongst predicted effects, and therefore PIGBLUP only calculates accuracies for traits, not indices. Accuracies are calculated whenever the EBV's are calculated. The accuracy calculation stages are apparent on the analysis progress screen.

The algorithm used for accuracy estimation is sensitive to pedigree errors in the data extract. These must be fixed for an analysis to run through successfully. To assist in this a new option has been provided in PIGCHECK to make detection and correction of pedigree errors easy and fast. A set of rules has been devised which allows PIGCHECK to build a file of possible corrections. This file may be displayed and edited and, if approved by the user, applied to the data extract. Corrections can include date-of-birth changes (based on the dates-of-birth of litter mates) so the extract should be re-sorted following correction and this is also possible within PIGCHECK.

On successful completion of an analysis, accuracies may be viewed/printed from the 'View EBVs' screen.

There are a number of new options allowing the display of EBV's Alone, EBV's and Accuracies together for each animal, and Accuracies Alone. When accuracies are to be displayed (either alone or together with EBV's), a set of sort options are displayed as well as a set of boxes which allow the display only of animals whose accuracies on particular traits lie within a nominated range.

Regarding sorting of accuracies, the direction of sorting may be chosen (descending accuracy is the default). Also, when both EBV's and Accuracies are to be displayed, sorting can be based on a trait's EBV or its Accuracy.

Selection Indexes

1. Background

Selection on an index simplifies the process of making selection decisions in a multiple trait situation. With an appropriately derived economic index, selection on that index will maximise the genetic progress in terms of profit.

PIGBLUP has included selection index calculations since its first release, either by simply applying a set of user-defined weights to the calculated EBVs (the uINDEX) or by applying a bio-economic model to the EBVs, making use of user-supplied inputs to combine the EBVs for average daily gain, backfat and number born alive (the $INDEX) (Stewart *et al*., 1990).

PIGBLUP V5.10 implements a third method based on work carried out by Neil Cameron when a visiting scientist at AGBU. A commercial pig unit is modelled following the proposal of de Vries (1989), with a group of gilts being bought in, mated a set number of times and then sold. The profit from this enterprise is calculated and compared with the profit after a trait has been improved by one unit, with all other variables fixed. The difference in profit is the economic value of that trait. Cameron and Crump (2001) reported economic values for average daily gain, backfat, number born alive in the first and later parities, feed conversion ratio and dressing percentage using the basic model which has now been implemented in PIGBLUP. Why implement a new approach? The new model will be easier to adapt for new traits in the future and currently handles more traits than the current $INDEX can.

It is only desirable to select on a single index within a line. However, PIGBLUP now supports the creation of multiple indexes from the EBVs. This functionality is intended to allow clients to assess different indexing possibilities, eg to assess, in their actual population, the effect of modifying the parameters defining an index. In addition this function will allow you to set up indexes relevant to any clients that you are marketing pigs to alongside your own selection index.

The multiple indexing in PIGBUP version 5.10 is available both at evaluation time and as part of the 'Buyer's Own Index' option..

2. Implementation in PIGBLUP 5.10

Existing PIGBLUP clients will notice the disappearance of the option to edit Economics. This has been incorporated in a new Indexes option that allows new indexes to be defined and existing ones to be edited. Note, defining or editing an index does NOT select it for analysis. The Setup/Parameters/Model option replaces the 'Use user-defined index' checkbox with an option for entering or picking operational indexes that are to be calculated during the Run Analysis stage. A stage has been added to the Analysis screen for each index selected.

Each breed sub-directory contains an indexes.txt file that lists all the currently defined indexes for that breed - whether operational or incomplete. The user gives each index a unique name and mnemonic. This index name is used with a type-specific extension (.eco, .wts or .mdl) to name a file containing each index's parameters.

The first step when creating a new index, is to input a breed-unique name for the index (without enclosed spaces) which will be used later when indexes are selected for analysis, a unique short mnemonic (eg $Inx) to be used as a column header for display and printing, and a longer description for the index.

Once these three fields have been completed, double clicking on the index name field reveals a radio box for index type selection.

The Indexes screen allows creation of three different types of index:
- A traditional PIGBLUP index (this is the previous PIGBLUP $Index) – saved with a .eco extension.
- A user-defined weights file (as of old except that the user can now opt for quadratic weights on particular traits) – saved with a .wts extension
- A model-derived index, based on Neil Cameron's work – saved with a .mdl extension.

Once the user has clicked one of the index type radio buttons, index-specific panels are displayed. PIGBLUP users will be familiar with the panels for the traditional PIGBLUP index and the (slightly enhanced) user-defined weights index. The model-derived index reveals a many-tabbed page control allowing for the alteration of the default settings for a target piggery. These inputs detail the size, economics and peculiarities of the target piggery that is to be modelled as well as the base levels for performance traits (ADG, BF, FCR, LMY, NBA1 and NBA2). A 'Check Valid' tab is provided that, when clicked, checks that all fields have been filled and contain 'sensible' values. If the current settings pass this validity check, a 'Test' tab is revealed which allows the user to experiment with EBV's to see if sensible indexes result.

All three index types are now checked for validity/sensibility. Whenever the list of defined indexes is displayed an icon is displayed to indicate whether PIGBLUP considers the index to be operational or incomplete. Only 'operational' indexes may be selected for analysis.

Should users choose to analyse multiple indexes, it is important to remember that the custom.inx file contains only the last-analysed index. The custom.inx file structure remains unchanged. Those users that back load the custom.inx file's index field into their herd-recording system can write a particular analysed index into the custom.inx file *post-analysis* using the Post-Analysis/Select Index for Upload to Herd Recording System option.

Cameron and Crump (2001) present economic values specific to certain conditions (eg an average carcase backfat of 18mm). The dependence of the economic values for a trait on the level of the trait itself can be accounted for by running a modified version of Neil Cameron's model specifically for animals rather than the whole group. This is how it has been implemented in PIGBLUP.

Selection and Mate Allocation

1. Background

Selection leads to genetic improvement, and selection on BLUP EBVs leads to more rapid genetic improvement than selection on other estimators of genetic merit. However, selection also reduces the genetic variation in the population, and this process is also faster with selection on BLUP EBVs. This is the result of selecting animals that are closely related – they have more genes in common and hence are less genetically variable.

Selection and mate allocation procedures have become more widespread in recent years, with the growth in the use of BLUP evaluations. They attempt to balance genetic progress and the degree of relationship among animals.

A detailed discussion of selection and mate allocation is beyond the scope of this paper but an example was discussed in the last Pig Genetics Workshop (Crump and Tier, 2001).

Dynamic procedures use information on the genetic merit (eg EBVs) and the relationships (eg inbreeding or a subset of the relationship matrix) for a specific set of candidate animals in order to decide which ones to select or which matings to allocate.

PIGBLUP has contained a dynamic mate allocation procedure since version 2.0 of the program (released in 1992). This procedure maximised the genetic merit of the offspring of matings while penalising inbreeding. Since this time a lot of research has gone on into dynamic selection and mate allocation procedures, and PIGBLUP now has a revised system.

Possibly the major limitation of the previous mate allocation procedure in PIGBLUP is that it only considered mate allocation. Selection decisions in pig populations are made much earlier than mates are allocated, and so the damage, in terms of increasing the relationships amongst candidates could already have been done.

2. Implementation in PIGBLUP 5.10

The new selection and mate allocation module in PIGBLUP (PBSAMA) is implemented as a DLL. This means that it is a library stored in a separate file called from the main program. This enables us to keep the PIGBLUP and PBSAMA functions as independent as possible, thereby avoiding conflicts and possibly simplifying some future upgrades.

Two closely related procedures are included in the PBSAMA module – selection and mate allocation. In each a list of candidate animals is processed to find the best combination of animals based on a merit function that can account for:

- Genetic merit,

- Relationships among the candidates,

- Relationships of the candidates with other animals that may join or remain in the breeding herd,

---

- Inbreeding of offspring (mate allocation only), and

- Variance of genetic merit (of offspring in mate allocation, of selected group in selection).

In order to take account of genetic relationships with other groups of animals that are in, or may enter, the breeding herd status information (in terms of a code, a date, a boar and a parity if appropriate) is required in the PIGBLUP file. Information on this will be circulated to herd management system developers and PIGBLUP clients. The codes indicate the current status of animals, eg pregnant, on-test, selected but not mature, and define the groups that relationships are considered with. For pregnant and lactating sows litters are inferred into groups (*in utero* or suckling). If status code information is missing, a reduced merit function is available (ie without relationships with other groups), but this is still a valuable tool. The user can set the weights applied to each part of the merit function, and any part of the merit function can be excluded.

In mate allocation minimum and maximum use of a boars can be set and any mating can be enforced or disallowed with the optimisation taking this into account.

The output from mate allocation consists of a table with candidate sows with a number of alternate boars chosen assuming that all other first-choice matings can be made. Selection output is a list of selected animals, each with a single alternate assuming that only that animal cannot be selected.

Acknowledgments

References

Crump, R. and B. Tier (2001) "Mate Selection" *2001 Pig Genetics Workshop*. pp 21-28.

Graser, H.-U. and B. Tier (1997) "Applying the concept of number of effective progeny to approximate accuracies of predictions derived from multiple trait analyses" *Proc. Assoc. Advmt. Anim. Breed. Genet.* Vol. 12, pp 547-551.

Stewart, T. S., D. H. Bache, D. L. Harris, M. E. Einstein, D. L. Lofgren and A. P. Schinckel (1990). "A bioeconomic profit function for swine production: application to developing optimal multitrait selection indexes." *Journal of Animal Breeding and Genetics* **107**: 340-350.