AN AUTOMATED ACROSS-HERD EVALUATION SERVICE FOR THE PIG INDUSTRY

G. M. Macbeth

Department of Primary Industries, Animal Research Institute, Yeerongpilly Q 4105

SUMMARY

The automation of BLUP evaluations through the National Pig Improvement Program (NPIP) makes across-herd selection possible for the entire Australian pig industry population. This level of automation requires a high degree of validation. The validation process used by NPIP is described.
Keywords: Pigs, genetic evaluation, national program

INTRODUCTION

In the Beef and Dairy industries annual across-herd determinations of breeding values have been sufficient for the identification and selection of superior performing breeder replacements. The pig industry is different, with shorter growing periods, pigs are selected on a weekly basis soon after their final weighing. This makes annual evaluations of limited value for on-going selection. Computer programs have been written to automatically analyse farm performance data and to produce timely reports on across-herd estimated breeding values (ebvs) from data sent by electronic mail. With data originating from many sources errors will occur and an automated data validation process is necessary.

MATERIALS AND METHODS

All data processing is on a 333Mhz Intel PC running Linux Redhat 4.2 with sequential linking of program modules organised by C-shell scripts. Two programs run in background (daemons) to automate data processing, one that retrieves and queues incoming data for analysis and the other to perform calculations and generate reports. The two daemons preserve the integrity of incoming data with that currently being analysed. Each of three breeds, Large White, Landrace and Duroc are analysed consecutively and preference for analysis is given to the breed waiting in queue the longest.

Electronic transfer validation. For each breed the last 10 years performance data is sent as an attachment to an email message. Processing of data is restricted to registered clients who are validated by their unique email address. Attachment file names must also correspond to a unique breed/herd name previously registered for each client.

General data validation. Each new data file is checked to see, if it is large enough to contain data (at least 200 lines), and if the file format is correct for reading. Although any data format can be accommodated, the standard format used for data retrieval is the PIGBLUP format (McSweeny 1998) which can be produced by several herd recording software packages. Next the new data is compared with old data received from previous evaluations. This is used to determine the level of discrepancies between the new and old data sets. Up to 250 modifications to the new data set are accepted with all changes recorded and stored for periodic review.

Animal pedigree validation. Unique across herd identifications are formed from a three-character herd prefix, which identifies each animal's birth place by herd and a unique within herd alpha-numeric code which does not change over time. Obvious combinations of pedigree errors are determined such as parents born before offspring, sire used as dam, male used as dam, female used as sire and animals with different parents. Some programming code logic is also used to determine which problem records are deleted prior to analysis.

Performance record validation. Data is accepted from a wide range of testing conditions and a large range in performance is accepted for analysis. This includes minimum and maximum days on test from 40 to 300 days, daily gain from 0.3 kg/day to 1.40 kg/day, backfat depth from 4mm to 31mm and a minimum end weight of 30 kg. Performance data falling outside the given ranges are deleted from the data set prior to analysis.

Contemporary group validation. Herd management groups are recorded in the on-farm recording software while within-herd contemporary groups are determined from final weighing dates using NPIP computer routines. The constraints for the formation of contemporary groups for production traits include, a minimum group size of 10 animals, at least two sires, and finish test span no more than 8 weeks. A preferred 4 week time span for grouping is set with the computer algorithm making several passes in an attempt to utilise as much data as possible given the constraints listed. Similarly contemporary groups are created for reproductive traits. The constraints used include, a minimum number of 15 sows, a maximum time span of 20 weeks, with a preferable time span of 6 weeks for grouping.

Herd linkage validation. Genetic linkage across herds, which is an essential component of across-herd comparisons, is assessed on the variance of estimated differences in genetic groups of base animals (Bunter *et al.* 1997). Herds that have sufficient links with other herds for any trait are pooled for across-herd analysis. Unlinked herds have ebvs determined and reported back to clients, but they will not have animals ranked across-herds as they cannot be compared with the same degree of accuracy.

Analysis. Breeding values are estimated by BLUP methodology and an animal model using PEST (Groeneveld 1990). The PEST Fortran source code was modified to create a separate file listing estimated breeding values (ebvs) for animal effect. This made it easy to retrieve animal ebvs for automated processing. The production trait model included, sex, within-herd contemporary group and herd management group as fixed effects, and litter effect as random with backfat corrected for weight at end of test. Litter size model included sow parity group, within-herd contemporary group, herd management group and service sire breed as fixed effects.

Post analysis validation. Quarterly log files of ebvs from all animals are stored for future detection of bias in genetic predictions such as the regression of recent (more accurate) on previous (less accurate) ebvs (Reverter *et al.* 1994). Finally, all report generating modules must be completed successfully prior to results being distributed by email. New data that has successfully been processed is archived for future validations of incoming data.

Reports generated. For each successfully completed evaluation, and for every herd sending data, the following reports are returned by email:

1) A list of ebvs (backfat, daily gain, number born alive and economic index) of all live animals for each producers herd.
2) a list of the latest across-herd evaluated ebvs of all AI boars available.
3) a list of ebvs, for all introduced animals, having progeny finishing test within the last 2 years.
4) a percentile report of ebv frequencies from all animals finishing test within the last 12 months.
5) a linkage summary detailing how well different traits are linked with other herds.

The total analysis and report generation process takes about 40 minutes for Duroc, 90 minutes for Landrace and 4 hours for Large White, with larger quantities of data taking longer to process. Currently there are 23,000 Duroc, 49,000 Landrace and 86,000 Large White animals with production records evaluated. If there are any problems with the processing of new data an email report, noting the source of the problem, is sent to the NPIP system administrator for immediate attention.

DISCUSSION

Automating data processing offers a considerable advantage over the previous 6 month batch processing offered by the National Pig Improvement Program (Macbeth 1999). Now producers have the ability to make selection decisions, not just by analysing their own data, but by using data pooled from a number of herds. The main benefit from national schemes is that progeny from imported animals (eg through AI) can be accessed more accurately compared to using only within herd performance information. This is more important in traits that have medium to low heritability such as growth rate (Macbeth 1997) where pooled information from other herds is more important.

Inevitable errors in data received from producers could easily bring a system to a halt while at the same time corrupting database records. The system has been designed to be robust and therefore a reliable provider of ebv estimates. During development Linux was also shown to be a very reliable yet cost effective operating system.

Resources are still required for continual monitoring of genetic evaluations. Quarterly log files of ebv estimates will be used to monitor changes in ebvs. Large changes may indicate incorrect formation of contemporary groups (Bertrand 1998), unaccounted fixed effects or inappropriate variance components (Reverter *et al.* 1994).

REFERENCES

McSweeny (1998) PIGBLUP users manual Version 4.0, Animal Genetics Breeding Unit, Appendix One, Interface Specifications, A1-A4

Bertrand, J.K. and Wiggans, G.R. (1998) Proc. 6th World Congr, Genet. App. Livest. Prod. **27**:425

Bunter, K.L. and Macbeth, G.M. (1997) Proc. Assoc. Advmt. Anim. Breed. Genet. **12**:607

Groeneveld, E. (1990) "PEST User's Manual". Institute of Animal Husbandry and Animal Behaviour, Federal Agricultural Research Centre, Germany

Macbeth, G.M. (1997) Qld. Pig Science Seminar, UQ Gatton College, 24 September 1997, p22-23

Macbeth, G.M. (1999) National Pig Improvement Program Genetic Evaluations ISSN 0727-6273

Reverter, A. Golden B.L., Bourdon, R.M. and Brinks, J.S. (1994) J. Anim. Sci **72**:34