Software applications for providing comprehensive computing capabilities to problems related to mixed models in animal breeding

Monchai DAUNGJINDA
(Department of Animal Science, Faculty of Agriculture, Khon Kaen University, Thailand 40002)

Abstract: Several computer packages have been developped to accomplish improved programs for animal breeding and genetic selection. This paper described most of the currant software and provided suggestions for improved software. Khon Kaen University, Thailand, will provide free of charge the new software developed at Khon Kaen University by the author of this paper. The contact for requesting the software is listed: moncha@kku.ac.th

Key words: mixed model; computer packages; animal breeding
CLC number: S813 Document code: A Article ID: 1001-411X(2005)S0-0078-10

Recently, several computer packages have been developed to accomplish problems related to mixed model in animal breeding. Special software for estimation of variance components and prediction of genetic merits are basically needed for genetic evaluation and selection program. Although there are some packages available on the internet, however, most of them are commercial or unfriendly to be used. The lists of recent software available on the internet are shown in Tab.1.

Most software is free license (mostly for academic purpose) with open-source code, however, UNIX is a main platform. To develop such program to windows platform requires specific compiler. Generally, data preparation and writing control file need particular management.

1 BLUPF90-PigPAK

The main objective of BLUPF90-PigPAK is to utilize models generally used in swine genetic evaluation with a user friendly graphic interface for PC and Windows users. Using powerful features from BLUPF90, PigPAK can perform a wide range of genetic evaluation functions. PigPAK can estimate variance components using REML and perform BLUP breeding value analysis from linear mixed models including random animal effects as additive genetic and permanent environment effect.

"BLUPF90-PigPack" is run using Microsoft EXCEL. It manages genetic evaluation using a wizard interface which allows you to create a VCE and BLUP report in five steps. Only performance data and pedigrees are required. It will automatically renumber the animals and count the number of effects in the model. Mate (or service sire) effects for reproductive evaluations and litter effects for growth evaluation can be included in data file with alpha-numeric format.

Biography: Monchai DAUNGJINDA, D.r., Professor, Department of Animal Science, Faculty of Agriculture, Khon Kaen University, Tailand, E-mail: moncha@kku.ac.th

Tab. 1 Recent software related to mixed model in animal breeding

name	details
ASREML	Statistical software designed for fitting mixed models for large datasets. Contains commented code for univariate and multivariate analysis including longitudinal and spatial analysis. http://uncronopio.org/asreml/homepage
DFREML	Variance component estimation software for animal breeding. Restricted maximum likelihood through a derivate-free method was developed by Karin Meyer. http://agbu.une.edu.au/~kmeyer/dfreml.html
BLUPF90 / REMLF90 / GIBBSF90	Animal breeding software package useful for animal genetic evaluation. Prediction of beeding value by BLUP and estimation of variance components by REML and Bayesian (Gibbs sampling) are all available. http://nce.ads.uga.edu/~ignacy/programs.html
PIGBLUP	BLUP software for genetic selection for swine. http://abri.une.edu.au
VCE/PEST	VCE software has been developed by Groeneveld for estimation of variance components. Document related to prediction of breeding value by PEST is also available. http://www.tzv.fal.de/~eg/
MTDFREML	Programs were developed by Keith Boldman and Dale Van Vleck. Debugging support have also been provided by Lisa Kriese and Curt Van Tassell. http://www.aipl.arsusda.gov/curtvt/mtdfreml.html

After variance component estimation or BLUP analysis is performed, PigPAK will create a BV report using the original animal ID. Users can keep data files and program files separately, however directories must be specified prior to analysis. BV reports with accuracy for up to four traits are available.

"BLUPF90" and related programs were developped in the lab of Dr. Ignacy Misztal with the purpose of providing comprehensive computing capabilities to problems related to mixed models in animal breeding. See http://nce.ads.uga.edu/~ignacy for details and documentation. These programs are mostly written in Fortran 90 and have a line-mode interface. "PigPAK" is a set of programs branched from "PCPAK" with graphical user interface for simple use in Windows. PigPAK was developped for the specific purpose of pig genetic evaluation, and parts of programs in the BLUPF90 family are available, which are:

BLUPF90: BLUP estimation using PCG method.

REMLF90: Variance component estimation using REML by EM algorithm.

AIREMLF90: Variance component estimation using REML by AI Algorithm.

RENUMMAT: Renumber program for creating data file and additive pedigree with animal ID in order number.

ACCF90: Approximate accuracy of BLUP solutions for direct maternal and multiple trait models.

All programs were compiled separately un-

der Microsoft Windows using Visual Fortran version 5.1. Users can run all programs separately using MS-DOS prompt or from the menu in PigPAK. This allows users to create all parameter files and BLUP reports with a point-and-click interface written in Visual BASIC for MS-EXCEL.

2 Specification

PigPAK requires Windows 95/98/ME/XP/2000 environment to install and Excel 98/2000/XP for running applications. It also requires at least 32 MB of memory and 5 MB of disk space for storing programs.

For BLUP and variance component estimation, BLUPF90 and REMLF90 support single and multiple trait models such as animal model, repeatability model, maternal dominance and random regression model with missing values accounted for and different models for each trait allowed. AIREMLF90 may not support some models and some particular data structures. ACCF90 is an approximation that works with repeatability and maternal models.

With the PigPAK wizard interface however, the single trait option will support only simple animal and animal with PE model. Multiple trait options will support up to four traits. Reports with accuracy are available only for single trait and multiple trait models. Variance component estimation by REML or AIREML is still available.

3 Model description

The following will provide more details of models that can be used in the analysis.

3.1 Basic animal model

PigPAK includes a basic animal model which allows animals in the data and animals in the pedigree to be included in the analysis so that all known relationships can be taken into account. Other effects, fixed and random, can be included for comprehensive use of mixed model technology. Fixed effects used in the model can be fitted as cross-classified variables and covariates. Combinations of fixed effects such as herd-year-season can be performed during the analysis, therefore, no additional data preparation is required. Normally, all traits analyzed with PigPack should be continuous rather than ordinal scale for proper use in the linear mixed model analysis. Models with a single record per animal such as growth traits can be analyzed using the basic animal model as follows:

$$y = X\beta + Za + \varepsilon, \text{ and } V\begin{bmatrix} a \\ \varepsilon \end{bmatrix} = \begin{bmatrix} A\sigma_\alpha^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix},$$

where y is vector of response variable, β is vector of fixed effects, a is vector of random additive genetic effects, ε is vector of random residual, X and Z are incident matrices related to fixed and random effects, A is numerator relationship matrix, σ_α^2 is additive genetic variance, and σ_e^2 is residual variance.

To perform BLUP, Henderson's MME can be written as:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+\alpha A^{-1} \end{bmatrix}\begin{bmatrix} \beta \\ a \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}, \quad \text{where} \alpha = \frac{\sigma_e^2}{\sigma_\alpha^2}.$$

3.2 Repeatability model

If multiple parity records are available, the permanent environment effect due to the animal needs to be taken into account. If mate effect from sire or litter effect for dam is available, it will be accounted as random effect as model described by Chen et al[1]. Fitting permanent environment effect as uncorrelated random effects is generally used in genetic evaluation. The

model for analysis is:

$$y = X\beta + Za + Wc + \varepsilon, \quad \text{and } V\begin{bmatrix} a \\ c \\ \varepsilon \end{bmatrix} = \begin{bmatrix} A\sigma_\alpha^2 & 0 & 0 \\ 0 & I\sigma_c^2 & 0 \\ 0 & 0 & I\sigma_e^2 \end{bmatrix},$$

where y is vector of response variable, β is vector of fixed effects, α is vector of random additive genetic effects, c is vector of random permanent environment effects, ε is vector of random residual, X, W and Z are incident matrices related to fixed and random effects, A is numerator relationship matrix, σ_a^2 is additive genetic variance, σ_c^2 is random permanent environment variance and σ_e^2 is residual variance.

To perform BLUP, Henderson's MME can be written as:

$$\begin{bmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z+\alpha A^{-1} & Z'W \\ W'X & W'Z & W'W+\nu I \end{bmatrix} \begin{bmatrix} \beta \\ a \\ c \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ W'y \end{bmatrix}, \quad \text{where } \alpha = \frac{\sigma_e^2}{\sigma_\alpha^2}, \ \nu = \frac{\sigma_e^2}{\sigma_c^2}.$$

3.3 Multi-trait model

PigPAK can also perform multi-trait analysis. Estimation of genetic correlations among traits and multivariate BLUP analysis can be accomplished as model used in NSIF[2] and NSR[3]. However, for graphic user interface, not greater than 4 traits are available. To perform beyond this, parameter editing is required and do the analysis from menu BLUP> Use old parameters. Multi-trait analysis can perform model with the same single records or same repeated records, and different model with single and repeated records. The following is bivariate model with the same single record:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix},$$

$$V\begin{bmatrix} a \\ e \end{bmatrix} = \begin{bmatrix} G \otimes A & 0 \\ 0 & R \otimes I \end{bmatrix},$$

where is vector of response variable for trait y_1 and y_2, β_1, β_2 is vector of fixed effects, α_1, α_2 is vector of random additive genetic effects, ε_1, ε_2 is vector of random residual, X, W and Z are incident matrices related to fixed and random effects, A is numerator relationship matrix, G is matrix of direct genetic variance-covariance for trait 1 and 2, R is matrix of residual variance-covariance for trait 1 and 2.

3.4 Sampling variance

If AREML option in PigPAK is performed, sampling variance for each variance components can be calculated. Therefore, standard error for heritability can be approximated using Lynch and Walsh[4] as follows:

$$\mathrm{Var}\left(\frac{\sigma_\alpha^2}{\sigma_t^2}\right) = \left(\frac{\sigma_\alpha^2}{\sigma_t^2}\right)^2 \left[\frac{\mathrm{Var}(\sigma_\alpha^2)}{(\sigma_\alpha^2)^2} + \frac{\mathrm{Var}(\sigma_t^2)}{(\sigma_t^2)^2} + \frac{2\,\mathrm{cov}(\sigma_\alpha^2, \sigma_t^2)}{(\sigma_\alpha^2)(\sigma_t^2)}\right]$$

$$\text{Then,} \quad SE(h^2) = \sqrt{\mathrm{Var}\left(\frac{\sigma_\alpha^2}{\sigma_t^2}\right)}$$

In addition, for multitrait analysis, standard error for genetic correlation can be approximated as follows:

$$\mathrm{Var}(r) = (\hat{r}) = (r)^2\left[\frac{\mathrm{Var}(\sigma_i^2)}{4(\sigma_i^2)^2} + \frac{\mathrm{Var}(\sigma_j^2)}{4(\sigma_j^2)^2} + \frac{\mathrm{Var}(\sigma_{ij})}{4(\sigma_{ij})^2} + \frac{2\,\mathrm{cov}(\sigma_i^2, \sigma_j^2)}{4(\sigma_i^2)(\sigma_j^2)} + \frac{2\,\mathrm{cov}(\sigma_i^2, \sigma_{ij})}{2(\sigma_i^2)(\sigma_{ij})} + \frac{2\,\mathrm{cov}(\sigma_{ij}, \sigma_j^2)}{2(\sigma_{ij})(\sigma_j^2)}\right]$$

Then,

$$SE(r) = \sqrt{\mathrm{Var}(r)}$$

3.5 How does program function

After installation, programs are stored in main or user-specified directory, ie C:/BLUPF90-PP. Two sub-directories of examples and helps are also created. Each directory will find the following programs and files Tab.2—Tab.4

Tab. 2 Main directory

name	type	description
PigPAK.xls	XLS	— Main graphic user interface — Creating BLUP and REML parameter files — Creating BLUP and VCE report in excel format — Creating genetic trend — Computing BV and h^2 for test day model
BLUPF90.EXE	PROG	— Computing BLUP solutions
REMLF90.EXE	PROG	— Estimating variance components using REML with EM algorithm
AIREMLF90.EXE	PROG	— Estimating variance components using REML with AI algorithm
ACCF90.EXE	PROG	— Computing approximate accuracy for BLUP solutions
RENUMMAT.EXE	PROG	— Renumbering animal in data and pedigree file in consecutive order number

Tab. 3 Examples directory

name	type	description
LWDAT.PRN	TXT	— Data file for model analysis with single trait and multi— trait with repeated records
LWPED.PRN	TXT	— Pedigree file for analysis with LWDAT
LWDAT.FMT	TXT	— Describe column number format for LWDAT.PRN

Tab. 4 Helps directory

name	type	description
WHOSWHO.TXT	TXT	— Accredit for key persons involved in BLUPF90 family
Manual—PP.PDF	PDF	— Manual for BLUPF90 PigPAK

When performed the analysis with wizard interface in PigPAK, All parameters entered in the form will be kept in particular Excel sheets. They will be written with corrected format as parameter file for RENUM and BLUPF90 using visual basic. Batch file to call the program with the parameter is also need to be created. VB in Excel has specific function to operate EXE file in this batch without closing the Excel program. Pedigree and solutions from the analysis are read to Excel sheets to join back the original animal ID, and also genetic trend will be created using chart function if user required.

Data and Pedigree files used in PigPack must be ASCII or TEXT file. Data must be in number, except for animal ID, Sire, Dam and Mate or Litter effect that can be alpha-numeric format. If create from Excel, save as PRN file (text file delimited with space) is preferable than Tab delimited or comma delimited.

If analyzing data is kept in different directory of programs. All execute programs will be copied to the data directory for simpler operation. After analysis a few file will be created, which can be copied to new name if need.

Suppose the original files for the analysis are LWDAT.PRN and LWPED.PRN, some additional files (Tab. 5) after analysis might be useful for later analysis, some additional files.

The default for convergence is 1d—08, however, users can choose their own by selecting menu.

4 Conditions of use

BLUPF90-PigPAK is distributed free of charge for academic and scientific use under the condition that it remains copyrighted. The use of any applications and compiled programs from PigPAK must be credited in any publications derived from their usage. For commercial or grant projects, personal communication of further agreements is required with any of the authors. There is no guarantee for correctness and there is no service for user purpose. However, specific

Tab. 5 Some additional files

name	type	description
RELWDAT. PRN	TXT	— Renumbered data file.
RELWPED. PRN	TXT	— Renumbered Pedigree file.
RENUM. PAR	TXT	— Parameter file for program RENUMMAT.
RENUM. MSG	TXT	— Log file from renumbering. Describe the levels of fixed and random effects after renum.
RENUM. PRN	TXT	— The details from renumbering. Describe how effects are combined and replications of each effects.
BLUP. PAR	TXT	— Parameter files for REML and BLUP analysis.
SOLUTIONS	TXT	— BLUP solutions file.
SOLUTIONS_VCE	TXT	— This file keeps variance estimation if performed.

questions, constructive criticism and debug reports are invited. Please email to monchai@kku.ac.th or ignacy@uga.edu.

BLUPF90-PigPAK has been made available on CD. However, the updated version is available at BLUPF90 homepage at KKU, http://agserver.kku.ac.th/monchai. The complete package provides program files, manuals and examples. Online registration is requested for further breeding and genetic group connection and to update version information. The registration page for BLUPF90-PigPAK is also available at BLUPF90 homepage.

References:

[1] CHEN P, BASS T J, Mabry JW, et al. Genetic correlations between lean growth and letter traits in US Yoikshire, Duroc, Hampshire, and Landrace pigs[J]. J Anim Sci, 2003, 81: 1700—1705.

[2] NSIF. Guidelines for Uniform Swine Improvement Programs[M]. Washington, DC: USDA, 1997.

[3] NSR. STAGES National Genetic Evaluation Trait Leader List[M]. IN: National Swine Registry, 2000.

[4] LYNCH M, WALSH B. Genetics and Analysis of Quantitative Traits[M]. CA: Sinauer Association, 1998.

(Edited by ZHOU Zhi-hong)

译文

动物育种中解决混合模型复杂计算问题的软件应用

Monchai DAUNGJINDA

(Department of Animal Science, Faculty of Agriculture, Khon Kaen University, Thailand, 40002)

摘要: 在动物育种和遗传选择的改良计划中应用的软件有很多个. 本文描述了其中最常用的软件, 并提出了有关软件的改进建议. 本文作者供职于泰国的 Khon Kaen大学, 可以免费提供其新开发的软件, 需要该软件的请与作者联系: moncha@kku.ac.th.

关键词: 混合模型; 计算软件; 动物育种

最近, 有数个软件包可以用于动物育种的混合模型分析. 对于遗传评估和选择方案来说, 进行方差组分估计和遗传性能预测需要特殊的软件. 虽然有些可以通过国际互联网下载, 可是, 多数这样的软件