MULTIVARIATE GENETIC EVALUATION IN SWINE
ON THE BASIS OF DATA FROM DIFFERENT TESTING SCHEMES

Milena Kovac and E. Groeneveld
University of Illinois, Dept. of Anim. Sci., Urbana 61801

SUMMARY
A computational strategy is presented that allows rapid implementation of genetic evaluations using multivariate mixed models. Data generated in different testing programs like field test of boars and gilts, litter recording schemes, and station test of sibs, may be combined to provide an estimate of the aggregate genotype. Pseudo code is given for the implementation of a "generic" testing structure illustrated by an example based on two traits from field test of boars and four traits from station test of sibs. BLUP's for all six traits are calculated for boars, parents and sib groups. Aggregate genotypes which correspond to the selection indices commonly used are calculated for selection candidates.

INTRODUCTION
Breeding objectives in pigs are complex comprising often growth, carcass composition, reproduction, and meat quality. This is reflected in multi trait selection indices which are widespreadly used to provide a prediction of the aggregate genotype. Boars and gilts are selected on the basis of data collected in a variety of testing schemes: performance records of the candidates, growth, feed consumption, and slaughter data from sibs collected at test stations, and fertility records from litter recording schemes. While BLUP has become the method of choice in dairy breeding its impact on swine breeding - so far - has been rather limited because of a serious discrepancy between requirements and models being used so far. The intent of this contribution is to present a multivariate model to predict the complex aggregate genotypes used in swine breeding. For compressed representation a multivariate genetic evaluation in swine is presented based on the boars own performance and data on relatives from test stations.

A GENERIC BREEDING PROGRAM IN SWINE
Traits and their sources to be investigated are derived from a 'generic' breeding program in swine. On farm performance test of boars or gilts (PT) is a standard procedure. Traits measured are usually live weight and ultrasonic or probed backfat Fixed effects to be considered may be breed, year/month of test, and a covariable live weight. Animal, litter and herd may be considered random effects. The structure of this type of data is straightforward with only one set of measurements for each boar or gilt.

Station test of sibs (SIB) is performed to assess growth potential and to collect carcass data that require slaughter of animals. Animals are often kept in groups which results in some of the measurement being made on each animal (e.g. daily gain, lean to fat ratio, and meat quality) while others are collected on the complete group (feed consumption). The effects to be considered may vary from one trait to the other. Fixed effects common to all traits are breed and test station. Daily gain and feed conversion will have to be adjusted for year/month of test. Meat quality parameters like pH or meat color are very sensitive to slaughter effects, requiring adjustments to each slaughter date. Lean:fat ratio requires adjustment for a common carcass weight and a seasonal effect like slaughter month/year. Data of a group are used to predict the genetic merit of related animals and not for the individuals themselves. It is sufficient to predict the genetic merit for the group. Averages for all traits could not be used, because sibs of

one group have different carcass weight and may be taken on different slaughter dates, placing the measurements into different classes of the fixed effect.

Thus, the models to be used are very complex with hardly the same incidence matrix for any two traits. Station test records are particularly complicated because of the combination of growth and slaughter data produced for either individuals under different environments or the sib group as a whole. Genetic groups have been included in this model because most likely every breeding program will introduce genetic material from outside populations. Residual covariance matrix **R** is a block diagonal matrix with two different submatrices: R_{01} - residual covariance matrix for PT traits and R_{02} for SIB traits. Additive genetic and residual covariance matrices in the purposed model are described in more details by Kovac and Groeneveld (1990).

COMPUTING ISSUES

The Mixed Model Equations (MME) can be setup in memory, preferably using sparse matrix storage, and solved either directly or iteratively, or the system may be solved by iteration on data. For ease explanation of the algorithm presented implies storage of coefficients in memory. Its basic structure follows the algorithm outlined by Groeneveld and Kovac (1990). The following description will point out the changes necessary to accommodate the more complicated structure in swine breeding data. The algorithm is split into two parts. Part 1 declares auxiliary variables that describe input and the model. This information is used in the second part which sets up and solve the MME.

PSEUDO CODE 1. DECLARATION OF AUXILIARY VARIABLES

```
 1 | parameters:                                    *** DECLARATION PART ***
 2 |    max_trait, max_effect = maximum number of traits or effects in a record
 3 | traits: 1- backfat (mm), 2- age (days), 3- daily gain (g), 4- meat quality(GOEFO),
 4 |    5- lean-to-fat ratio, 6- feed conversion
 5 | effects: 1- season (F), 2- herd (R), 3- station (F), 4- live weight (Cov), 5- carcass
 6 |    weight  (Cov), 6- slaughter date (fixed), 7- animal / sib group (R)
 7 | DATA_FILES
 8 | integer  what_trait (max_trait, #data_files), what_effect (max_effect, #data_files)
 9 | boar.data record contains effects: 1, 2, 7, and 4 and two traits: 1 and 2.
10 |    what_trait(1 to 7, 1)= 1,2, 5*none      what_effect (1 to 7, 1)=1, 2, 7, 4, 3*none
11 | in data_format(1) = '(i1,2i3,f6.0,t20,f5.0,t14,f6.0)'
12 | station.data contains seven columns for effects and seven measurements:
13 |    what_trait(1 to 7,2)=3, 4, 5, 3, 4, 5, 6   what_effect (1 to 7, 2) = 1, 3, 6, 7, 6, 5, 5
14 | in data_format(2) = '(i1, 2i3, t13, ........................ t33, f5.0)'
15 | pedigree.data record contains animal, sire, dam,
16 |    pedigree_type (1 = animal, 2 = sib group)   birth_date ( for inbreeding)
17 | in pedigree_format '(i1,3i3)'
18 | MODEL connects measurements and effects in a data record
19 | which_effect(max_trait, max_effect, #data_files)
20 | station.data: Effects for lean:fat ratio (trait 3) are placed in positions 1,2, and 4.
21 |    Covariables are in position 6 for animal one and 7 for animal two.
22 | ....which_effect (3,1 to 4, 2)=1, 2, 4, 6      which_effect (6,1 to 4, 2)= 1, 2, 4, 7
23 |    which_effect (7,1 to 4, 2)=1, 2, 4, 0
```

Declaration of auxiliary variables. In the declaration part (Pseudo Code 1) lines 3 through 6 define all traits and effects in the MME. A separate file is used for each test environment, i.e. "boar.data" for PT and "station.data" for SIB. File "boar.data" (lines 9-11) contains effects season, herd, animal, and covariable live weight described by codes 1,2,7,4 in *what_effect (?,1)*. Two traits - backfat and age - are measured. Their

codes 1 and 2 are stored in variable *what_trait(?,1)*. The other data files are described correspondingly. The "pedigree.data" (lines 15-17) contains animal - parents and sib group - parents relationships. To distinguish them, *pedigree_type* is added to each pedigree record. If inbreeding is to be included, birth date has to be provided as well. Furthermore, *which_effect* (lines 18-23) describes the model and connects measurements and effects in any data file. Its first index refers to all measurements collected on the same animal or sib group, while the second one denotes the effects. The specification of lean to fat ratio R from "station.data" will serve as an example. R (line 22) has effects season, herd and animal on positions 1, 2, 4 respectively. Furthermore, R has to be adjusted to an individual carcass weight, which is found on position 6 for the first animal - *which_effect(3,4,2)* and on position seven for the second animal - *which_effect(6,4,2)* among seven effects collected for station data.

PSEUDOCODE 2. SETTING UP AND SOLVING MIXED MODEL EQUATIONS

24	Count traits for each effect. Store results in traits_in. *** **ACTIVE PART** ***
25	Determine position for traits within effects (trait_pos).
26	Determine start for each effect as a function of #levels for effects and trait_in.
27	Compute diagonal elements of D^{-1} as needed for modified A^{-1}.
28	**Measurements from many data_files**
29	do what_file = 1, 2, ...#data_files
30	OPEN file with name stored in data_file(what_file). It must exist.
31	Read current_levels, covariables, and measurements until end_of file
32	Determine elements of incidence matrix **W** and addresses in the solution vector
33	do for _trait = 1, 2, ..., number of measurements in a record from what_file
34	do for_effect = 1, 2, ..., number of all effects in a record from what_file
35	current_level = CL = which_effect(i, j, what_file)
36	w_cov (for_trait, for_effect) = {1 or covariable(CL)}.
37	address(for_trait, for_effect) = function{start, nesting, CL, trait_in, trait_pos}
38	end do
39	Choose the appropriate residual covariance matrix (what_rvar).
40	Build MME using address, w_cov, measurements, and $R_0(1, 1,$ what_rvar)
41	Build Right_Hand_Side.
42	end when end_of_file is reached
43	end do if all data files are read
44	**Random effects**
45	do effect = 1, 2, ...all_effects
46	if relationship then call add_relationship and use D^{-1} and pedigree file
47	else add covariance matrix G_0 (effect) to the diagonal of the coefficient matrix
48	end do
49	**System of equations is set up. Solve it** by any suitable subroutine.
50	Compute agregate genotype of the candidates for selection.

Setting up the coefficient matrix. With the auxiliary variables defined as above the algorithm for setting up the coefficient matrix becomes very compact (Pseudo Code 2). The outer loop determines all files involved. All data files are read by the same subroutine called in line 31. To skip zero equations the addresses for the coefficient matrix calculated by a function called in line 37 requires variables *start, trait_in, trait_pos*, and *which_effect*. Making the appropriate additions to the coefficient matrix and the right hand side is then straightforward. Random effects are handled in lines 44 through 48. The elements of a diagonal matrix D^{-1} needed to set up modified numerator relationship matrix do not change during computations. Therefore, they may be calculated in the first round only (line 27), if the coefficient matrix has to be set up repeatedly.

DISCUSSION

Many well organized swine breeding programs use complex aggregate genotypes predicted by a relatively large number of traits measured in a number of different test environments (Christensen, et al., 1986 and Bichard, et al., 1986). They list sow productivity, economy of gain, and suitability of carcass as breeding objectives, being estimated on the basis of the candidate's own performance and performance of its relatives. Reported BLUP implementations, so far, are rather limited in their scope. Most of the models used are univariate (Hudson and Kennedy, 1985) or transformed multivariate (Schinckel, et al., 1986). Few multiple trait MME application have been reported combining traits with identical incidence matrices as Groeneveld and Kovac (1989) in the context of variance component estimation.

The procedure outlined allows implementation of multi trait genetic evaluation for nearly any combination of testing schemes in swine breeding. Data from testing schemes not considered in the example like litter records on gilts can be added to the system by specifying another data file. This allows users of complex selection indices to move to BLUP. Apart from its theoretical superiority there are substantial other benefits connected with genetic evaluation based on mixed model methodology. Because BLUP's for genetic merit are cumulative over time they facilitate monitoring of genetic change in the population permitting efficiency control of breeding programs. Across herd evaluation gives a direct comparison of the genetic level of each herd. This, in turn, allows to choose nucleus herds on the basis of their genetic merit. Across herd evaluation requires genetic ties which are provided by a certain portion of identical genes as they are created by exchange of related breeding stocks amongst herds.

Computational load is often cited as an impediment to the implementation of multi trait mixed model evaluations. It has to be acknowledged that BLUP requires more CPU time than calculation of a selection index. This is not a problem any longer considering the dramatic advance in computer hardware. Furthermore, solution vectors should be preset to the old solutions in routine application of iterative procedures. The new equation system expanded by animals to be evaluated for the first time will converge after a few rounds. The outlined procedure has been applied to multivariate genetic evaluation of 130,000 test station records to estimate the genetic trend. Each of the four traits had a different statistical model, i.e. incidence matrix. The system of around 800,000 equations was solved on a SUN 4 workstation using 24MB. Memory requirement of this magnitude seems to be high. It has, however, to be noted that with rapidly increasing memory size it is not efficient to use short cuts in software that restrict the software in its generality. Because these restrictions are part of the internal algorithm of the software they are difficult to overcome, usually only by rewriting the program. On the other hand, if the most comprehensive algorithm is implemented, it can, by definition also deal with simpler cases.

REFERENCES

Bichard, M., David, P. J. and Bovey, M. 1986. 3rd World Congress on Genetics applied to Livestock Production 10: 130-142.

Christensen, A., Sorensen, D. A., Vestergaard, T. and Van Kemenade, P. 1986. 3rd World Congress on Genetics applied to Livestock Production 10: 143-148.

Groeneveld, E. and Kovac, M. 1989. J. Dairy Sci. submitted.

Groeneveld, E. and Kovac, M. 1990. J. Dairy Sci. 73: in print.

Hudson, G. F. S. and Kennedy, B. W. 1985. J. Anim. Sci. 61(1): 83-91.

Kovac, M. and Groeneveld, E. 1990. J. Anim. Sci. submitted.

Schinckel, A. P., Harris, D. L., Stewart, T. S. and Lofgen, D. L. 1986. 3rd World Congress on Genetics Applied to Livestock Production 10: 98-109.