United States
Department of
Agriculture

**Agricultural
Research
Service**

June 1995

A Manual for Use of MTGSAM

A Set of Fortran Programs to Apply Gibbs Sampling to Animal Models for Variance Component Estimation

Draft

Van Tassell, C. P. and L. D. Van Vleck. 1995. A Manual for Use of MTGSAM. A Set of Fortran Programs to Apply Gibbs Sampling to Animal Models for Variance Component Estimation[DRAFT]. U.S. Department of Agriculture, Agricultural Research Service.

The manual in five chapters describes how to compile and run the Fortran programs for multiple trait Gibbs sampling for estimation of variances components. Chapters describe various models that can be used, description of computing strategies, the theory behind the method.

Note: This manual is a draft version of the technical documentation that will accompany the MTGSAM. It is being released with the software on an interim basis until the official documentation has been prepared as a publication of the U.S. Department of Agriculture's Agricultural Research Service.

Van Tassell and Van Vleck are with the Roman L. Hruska U.S. Meat Animal Research Center, Agricultural Research Service, U.S. Department of Agriculture (A218 Animal Sciences, University of Nebraska, Lincoln, NE 68583-0908).

Introduction

Gibbs sampling (GS) is a method of numerical integration that allows inferences to be made about joint or marginal densities, even when those densities cannot be evaluated directly. The GS algorithm is based on generation, in sequence, of variables from all of the full conditional densities. The full conditional density is the density of a variable given all other parameters in the model. For example, if GS is used to estimate the distributions of f(a|y), f(b|y), or f(a,b|y), then the full conditional distributions, f(a|b,y) and f(b|a,y), would be required. In order to use GS to evaluate any of these densities, an arbitrary starting value for one of the variables would be chosen, and then values would be drawn from the full conditional densities in the sequence

$$a^n \sim f(a|b^{n-1},y) \text{ and}$$

$$b^n \sim f(b|a^n,y),$$

where ~ indicates that the variable is a random variable from the distribution specified, and the superscript refers to the sequence of the value in the GS chain. If the sequence is repeated enough times, the distribution of the a and b samples will be from the distributions f(a|y) and f(b|y), and the a,b sample pairs will be drawn from the f(a,b|y) distribution.

In the case of the problem of estimation of VC, the joint density of interest is the distribution of fixed effects, random effects, and VC, all given the data. The marginal densities of interest in this problem are the distributions of fixed effects, random effects, or VC, given the data.

A general set of Fortran programs was developed for estimation of variance components with animal models using GS. The programs are called **M**ultiple **T**rait **G**ibbs **S**ampling in **A**nimal **M**odels (**MTGSAM**). The program interface is similar to that used in the MTDFREML programs (Boldman et al., 1993) and shares a substantial amount of Fortran programs. The programs support multiple-trait models with an arbitrary number of covariates, fixed effects, second animal genetic effects, and random effects for each trait. The programs manage data with any combination of missing observations. Genetic effects and uncorrelated random effects related to animal or second animal effects (e.g., direct or maternal permanent environmental effects) are generated as a block to increase the rate of convergence. The programs generate means and

samples from the GS chain for estimating variance components, variance ratios, heritabilities, fixed and random effects, and contrasts. In order to guarantee valid parameter estimates, proper Bayesian prior distributions may be required for variance components. The programs generate the Gibbs samples but will not perform burn-in or convergence analysis and will not generate posterior distributions. Therefore, the user is expected to have an understanding of these aspects of Gibbs sampling. Some of these features may be added over time, as better algorithms are developed.

Chapter 1 gives instructions about how to compile and execute the programs. Chapter 2 presents some small numerical examples. Chapter 3 presents some of the theoretic results related to Gibbs sampling. Chapter 4 discusses some of the computational strategies employed to help if modifications of the programs are needed.

MTGSAM Fortran Files

MTGSNRM This program computes $\mathbf{A}^{-1}$ and recodes identification numbers for animal, sire, and dam. This program is nearly identical to MTDFNRM, the analogous program used by MTDFREML. The program uses subroutines in MTGSSUB.

MTGSPREP This program allows the user to specify the traits to be analyzed and to specify the model for each trait. The program then reads the original data file and generates a new data file with recoded levels of fixed and random effects, and information used to build the mixed model equations. Note that any animal that appears in the data file must be included in the pedigree supplied to MTGSNRM. This program is similar to MTDFPREP, the analogous program used by MTDFREML. The program uses subroutines in MTGSSUB.

MTGSRUN This program generates Bayesian posterior distributions and means using a Gibbs sampling algorithm. Means for variance components, fixed and random effects, heritabilities, correlations, and contrasts can be calculated. The program generates files containing Gibbs samples with a user specified burn-in period and interval between samples. The program uses subroutines in MTGSSUB and MTGSRSB.

MTGSSUB A set of general subroutines used by all of the main programs.

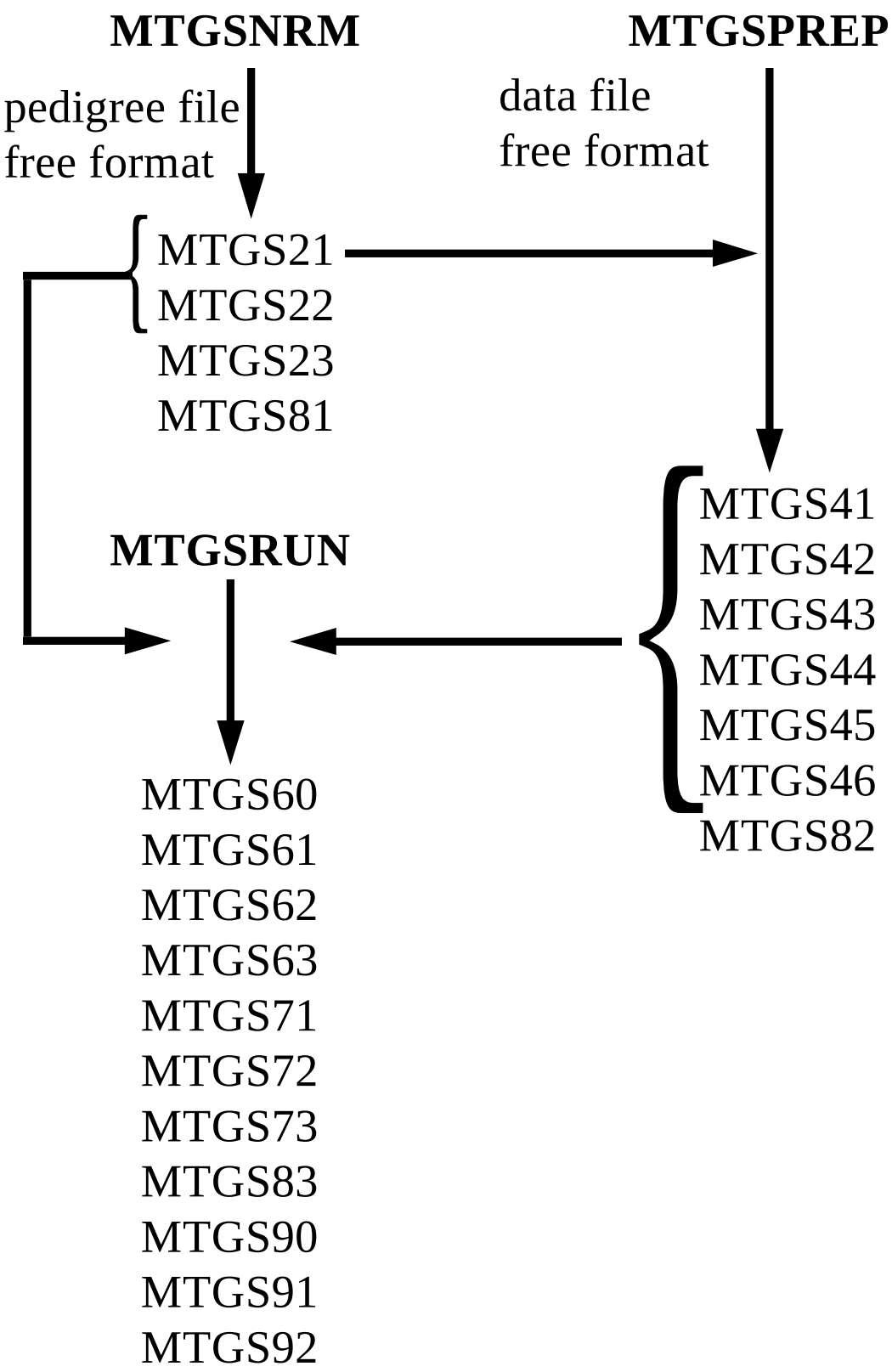
MTGSRSB A set of subroutines used only by MTGSRUN.

MTGSAM Data Files

File	Format[1]	Description	MTGSNRM[2]	MTGSPREP[2]	MTGSRUN[2]
MTGS21	U	reordered IDs	O	I	I
MTGS22	U	non-zero $\mathbf{A}^{-1}$ elements	O	-	I
MTGS23	F	inbreeding information	O	-	-
MTGS41	F	model information	-	O	I
MTGS42	U	recoded data	-	O	I
MTGS43	U	animal summary	-	O	I
MTGS44	U	blocking information	-	O	I
MTGS45	F	covariate and fixed level labels	-	{O}	{I}
MTGS46	F	uncorrelated random level labels	-	{O}	{I}
MTGS60	F	rename for input file for new run	-	-	{I/}O
MTGS61	U,D	samples from GS chain	-	-	O
MTGS62	U,D	parameters of GS densities	-	-	O
MTGS63	F	data needed to analyze 61 & 62	-	-	O
MTGS71	F	posterior means of covariates and fixed effects	-	-	{O}
MTGS72	F	posterior means of animal effects	-	-	{O}
MTGS73	F	posterior means of uncorrelated random effects	-	-	{O}
MTGS81	F	log file for MTGSNRM	O	-	-
MTGS82	F	log file for MTGSPREP	-	O	-
MTGS83	F	log file for MTGSRUN	-	-	O
MTGS90	U	checkpoint file	-	-	{I/}O
MTGS91	U	checkpoint file	-	-	{I/}O
MTGS92	U	checkpoint file	-	-	{I/}O
MTGS99	U	scratch file	-	-	I/O/D

[1]File format: F = formatted, U = unformatted, D = direct access

[2]File type: I = input, O = output, D = deleted at end of run, {} indicates that the use of this file varies with the specific run.


MTGSNRM
MTGSPREP
pedigree file
free format
data file
free format
MTGS21
MTGS22
MTGS23
MTGS81
MTGS41
MTGS42
MTGS43
MTGS44
MTGS45
MTGS46
MTGS82
MTGSRUN
MTGS60
MTGS61
MTGS62
MTGS63
MTGS71
MTGS72
MTGS73
MTGS83
MTGS90
MTGS91
MTGS92

CHAPTER ONE: User Notes for MTGSAM

Introduction

Multiple **T**rait **G**ibbs **S**ampling for **A**nimal **M**odels, denoted as MTGSAM, is a set of programs to estimate Bayesian posterior means and distributions for (co)variance components, fixed and random effects, and contrasts using Gibbs sampling with animal models. The interface of the programs are similar to the MTDFREML (**M**ultiple **T**rait **D**erivative-**F**ree **Re**stricted **M**aximum **L**ikelihood) programs. These two programs share a significant amount of Fortran source code, especially for the numerator relationship and data manipulation programs.

The MTGSAM programs can be used for a single trait analysis or for any number of traits. The programs allow any combination of missing observations for the multiple trait models. Posterior mean estimates for "fixed" and random effects and for contrasts can also be calculated. The size of analyses that can be run depends on the number of traits and animals in the analysis, and on computer speed and memory.

Animal models can incorporate additive genetic effects not only for animals with records, but also for parents and other relatives without records included in the pedigree file. One additional correlated random effect, (e.g., maternal genetic) and any number of uncorrelated random effects can be used for each trait in the analysis. Fixed effects and covariates are specified separately for each trait.

Computing Considerations

All programs are written in FORTRAN 77. The programs were developed on 486 and Pentium microcomputers using a Microsoft Fortran Powerstation compiler, but should run with minor modifications on any platform with a FORTRAN compiler. At least 16 MB of memory is advisable, especially for the MTGSRUN program. Requirements for hard disk space will vary with the amount of information requested in MTGSRUN. If only posterior means are requested and no sampling information is written, then the disk requirements are minimal. If the number of

Gibbs samples or elements written in each sample increase, the disk space will increase as well. If only variance component information is written, and the number of Gibbs samples is relatively small, several MB of disk space should be sufficient. If the "solutions" are written in each Gibbs sample or the number of samples written is relatively large, the amount of disk space can be quite large (hundreds of MB).

Currently, all programs have interactive input/output defined to standard FORTRAN 77 units of 5 (input from keyboard) and 6 (output to screen). Two areas that may need modification on platforms other than PC's are the input/output file connections and the timing routines. Currently, the Fortran OPEN statement is used to open files and connect to the appropriate unit numbers. The files used by the program all have names of the type MTGS##, where ## corresponds to the unit number of the appropriate MTGSAM program. For example, IUN81 (unit 81 in MTGSNRM) corresponds to file MTGS81. The compiler or system dependent timing subroutines are in the file MSTIME. This file contains two subroutines: DOSTIM and DOSDAT, which return the system time and date, respectively. These calls correspond to the Microway NDP Fortran compiler calls for timings and MSTIME should be eliminated when using that compiler. These timing subroutines can be changed for other compilers or platforms, or commented out (C in column 1 of the source code).

The file GSPARAM.FOR contains maximum parameter definitions for arrays in programs MTGSPREP and MTGSRUN. The INCLUDE statement brings GSPARAM.FOR into MTGSPREP and MTGSRUN to provide consistent parameter definitions. In the PC environment, the INCLUDE statement looks like :

```
INCLUDE 'GSPARAM.FOR'
```

In MVS/TSO environments, the INCLUDE statement is :

```
INCLUDE 'QCAROL.GSPARAM.FOR'
```

In CMS environments, a library must be created that contains the INCLUDE source code. A possible exec to create the library containing the INCLUDE source code is:

```
/* An exec to use INCLUDE statement in MTGSAM  */
TRACE RESULTS
"MAC GEN LIB1 GSPARAM"
"GLOBAL MACLIB LIB1"
```

The GSPARAM file *must* have a filetype of COPY in the CMS environment. After the library is created on CMS, the INCLUDE statement in MTGSPREP or MTGSRUN has the form:

INCLUDE(GSPARAM)

Check the system documentation for variations of this code if problems arise. Alternatively, the source code in GSPARAM.FOR could be placed directly into MTGSPREP and MTGSRUN programs replacing the INCLUDE statement. Note that if GSPARAM.FOR is changed, both MTGSPREP and MTGSRUN must be recompiled, i.e., the programs must have the same parameter definitions.

MTGSAM Programs

The three MTGSAM programs: 1) form the inverse of the relationship matrix, 2) prepare data to set up of the weighted least squares part of MME and 3) set up and solve the MME for solutions for fixed and random effects using Gauss-Seidel iteration (**GSI**) and/or estimate posterior means and generate Gibbs samples for (co)variance components, fixed and random effects, contrasts of fixed and random effects. The following programs and subroutines are needed:

MTGSNRM	Forms non-zero elements of A^{-1} using an ASCII free formatted pedigree file using the rules of Quaas (1976). Program reorders animal, sire and dam identification (for animal model) or alternatively, sire, sire of sire and maternal grandsire of sire identification (for sire models). The program also generates inbreeding coefficients for animals, sires and dams.
MTGSPREP	From an ASCII data file, forms coefficients for MME according to model specifications supplied to the program.
MTGSRUN	From coefficients of MME formed in MTGSPREP there are two options in MTGSRUN. First the program can be used to obtain solutions to the MME using user supplied values for (co)variance components using Gauss-Seidel iteration. The second option is to run some rounds of Gauss-Seidel iteration followed by Gibbs sampling. The Gibbs sampler can be used to obtain Bayesian estimates of posterior means of (co)variance components, heritabilities, and correlations; covariates, fixed, and random effects; and contrasts. The Gibbs samples may also be saved to files

	to allow construction of univariate or multivariate posterior density estimation.
GSPARAM.FOR	Fortran code for INCLUDE statement that contains parameter statements for maximum limits for variables such as maximum number of animals and fixed effects. If the user chooses not to use the INCLUDE statement, these source statements can be placed directly into source code for MTGSPREP and MTGSRUN where the INCLUDE statement is located.
MTGSSUB	A file containing a suite of general purpose Fortran subroutines needed in the MTGSAM programs. Routines were written by C. Van Tassell, K. Meyer (1988), K. Boldman (Boldman et al., 1993), and P. VanRaden. Some subroutines were also obtained (and some modified) from STATLIB and NETLIB.
MTGSRSB	A set of subroutines used only by MTGSRUN. Subroutines written by C. Van Tassell.

Setting up A^{-1} (MTGSNRM)

Pedigree information may be in a file separate from the data when animals without records need to be included in the relationship matrix. The file is assumed to be in free format, i.e., all variables are separated by spaces. The source code can be easily modified to accommodate formatted read statements if animal identification numbers are not separated by spaces. The pedigree file needs to include numeric fields for:

- Animal ID, sire ID and dam ID (optionally sire ID, sire of sire ID and maternal grandsire of sire ID for sire models).
- If both parents of an animal *without* a record are unknown, that animal does not need to be in the pedigree file as an animal because it does not provide ties.
- If a sire ID or dam ID is missing, the missing parent ID must be coded as a 0.
- If an animal has missing parent(s), and if the missing parent is needed to code for a maternal or paternal correlated random effect, the missing parent must be coded with a unique number other than 0.
- The largest ID the program can accommodate is the maximum integer the compiler can handle ($2^{31} - 1 = 2{,}147{,}483{,}647$ for most Fortran compilers). If IDs are larger than this or include characters, IDs must be recoded prior to running MTGSNRM.

In the parameter statement of MTGSNRM, the maximum number of animals (MAXAN) in the relationship matrix can be changed. MAXAN represents both animals with records and base animals.

To Run MTGSNRM:

1. Compile and link MTGSNRM and MTGSSUB (subroutines).
2. Run MTGSNRM. This program will calculate A^{-1}. The program asks :
    a) Do you want to calculate A^{-1} using animal, sire, and dam (0) or sire, sire of sire, and maternal grandsire of sire (1) rules?
    b) Maximum animal ID in pedigree file (for data verification only).
    c) Minimum animal ID in pedigree file (for data verification only, 0 will work).
    d) Name of free formatted file containing pedigree information, e.g., ANIMAL.PED.
    e) Number of integer fields in pedigree file.
    f) Position of animal (or sire) ID in vector of integers.
    g) Position of sire (or sire of sire) ID in vector of integers.
    h) Position of dam (or MGS of sire) ID in vector of integers.

Note that the columns specifed in questions g and h can correspond to a vector of zeros if one choosed to ignore relationships among animals. The number of animals in A^{-1} will appear on the screen and in file MTGS81. This number is needed later for MTGSPREP - write it down.

The MTGSNRM program produces three output files :

MTGS21 This file contains the number of animals, followed by vector of original animal IDs sorted in ascending order in binary format. The location in the file corresponds to the recoded ID, i.e., position i is the original ID corresponding to recoded ID i.

MTGS22 This file contains the non-zero lower-half-stored elements of coefficients of A^{-1} in binary format. Note that the elements are NOT summed. The format of the file is i, j, $a^{i,j}$.

MTGS23 This file contains the number of animals, followed by one line for each animal in the pedigree in the following format:

Recoded ID			Original ID			Inbreeding Coefficient		
Animal	Sire	Dam	Animal	Sire	Dam	Animal	Sire	Dam

MTGS81 This is the log file of information from the execution of MTGSNRM. The information includes number of animals in A^{-1}, number of non-zero elements, and inbreeding information.

Setting up MME (MTGSPREP)

The data file for MTGSPREP must have integer variables first, including animal ID, numerical identities for fixed effect levels, and numerical identities for random effects (e.g., dam ID for maternal or permanent environment effects), followed by real variables, which include covariates and trait measurements. The data file should be in free format with at least one space separating variables. If the data is not in free format, the source code can be modified for a formatted read of unit IUN31.

Models can be different for each trait in the analysis. The number of fixed effects, covariates or uncorrelated random effects are usually not limiting for a trait. An INCLUDE file, GSPARAM.FOR, contains maximums for several variables used in the programs. The limits must be large enough to accommodate the data set. If not, error messages or wrong results will be obtained. Limits that can be changed in GSPARAM.FOR are :

MAXTRT	maximum number of traits in the analysis
MAXINTR	maximum number of integer variables on each record
MAXR8	maximum number of real variables on each record
MAXCOM	maximum lines of comments for output description of analysis
MAXCOV	maximum number of covariates per trait
MAXNFR	maximum number of regression coefficients per trait
MAXFIX	maximum number of fixed effects per trait
MAXNFL	maximum number of levels for each fixed effect
MAXANIM	maximum number of animals
MAXRAN	maximum number of uncorrelated random effects per trait
MAXNRL	maximum number of levels for each uncorrelated random effect
MAXCNT	maximum number of contrasts to monitor
MAXCCOEF	maximum number of contrast coefficients for each contrast
NZEC	maximum number of non-zero elements in half-stored coefficient matrix, note that $\mathbf{A}^{-1}$ is not added to MME
NZEA	maximum number of non-zero elements in **full**-stored $\mathbf{A}^{-1}$ matrix
NZE	larger of NZEC or NZEA - usually NZEC

Fields in the data file can be used for more than one trait and can have more than one name within or across traits. For example, for weaning weight in beef cattle, when additive, maternal and permanent environmental random effects are in the model, the dam ID field can be used to indicate both maternal genetic and permanent environment effects. More than one uncorrelated random effect can be specified for each trait in the analysis. Within trait,

uncorrelated random effects will, of course, be uncorrelated. However, if the same uncorrelated random factor (i.e., in the same column in the data set) is used across traits, a covariance can be estimated. Uncorrelated factors which occur in the same column across traits are considered to be in the same uncorrelated random "group," i.e., factors coded in the same column may have non-zero correlations, but groups coded in different columns are assumed to have zero correlations.

The MME set up by MTGSPREP have the following order :

covariate(s) trait 1
⋮
covariate(s) trait n
fixed effect(s) trait 1
⋮
fixed effect(s) trait n
additive genetic animal effect trait 1
⋮
additive genetic animal effect trait n
additional correlated random effect trait 1 (e.g., maternal genetic)
⋮
additional correlated random effect trait n
uncorrelated random effect(s) for trait 1
⋮
uncorrelated random effect(s) for trait n

The number and types of equations in the MME depend on specific models and data. Equations in the above list that do not apply to a specific analysis do not appear. All models will have additive genetic animal effects. Uncorrelated animal effects will result from a pedigree file with all sires and dams missing. Genetic variances cannot be estimated if A = I for an animal model although a sire model can be used with A = I.

To run MTGSPREP

1. Compile and link MTGSPREP and MTGSSUB (subroutines). GSPARAM.FOR must be available.
2. Run MTGSPREP. The program reads MTGS21 and asks the following questions:
   a) Name of data file (IUN31), e.g., ANIMAL.DAT
   b) Description of analysis (up to 6 lines, terminated with a * in column 1 after last comment line)
   c) Number of integer variables in each line of data file
   d) Number of real variables in each line of data file

e) Number of traits in the analysis

Questions f-y are repeated for each trait with the exceptions of q) and r)

f) Name of trait
g) Position for trait in list of real variables
h) Missing value designation for trait (e.g., 0,0.0, -999.9, etc.)
i) Number of covariates

Questions j-l will be repeated for each covariate in a trait

j) Name for first covariate
k) Position of first covariate in list of real variables
l) Type of covariate (linear, quadratic, etc)

m) Number of fixed effects

Questions n-p will be repeated for each fixed effect in a trait

n) Name of fixed effect
o) Position of fixed effect in list of integer variables
p) Write levels of fixed effect to unit 82 (MTGS82): 1 yes; 0 no

Questions q-r will be asked only for the first trait

q) Position of animal ID in list of integers (same for each trait)
r) Number of animals in A^{-1} (from MTGSNRM)

s) Is there a second animal (e.g., maternal genetic) effect (1 yes; 0 no)

If there is a second animal effect for the trait, answer questions t) and u)

t) Name of second animal effect
u) Position of second animal effect in list of integer variables

v) Number of uncorrelated random effects (e.g., PE, litter)

Questions w-y will be repeated for each uncorrelated random effect in a trait

w) Name of uncorrelated random effect
x) Position of uncorrelated random effect in list of integers
y) Write levels of uncorrelated random effects to unit 82 (1 yes; 0 no)

Question z will be asked if there is at least one covariate or fixed effect

z) Save original labels to match with mean estimates for covariates and fixed effects in MTGSRUN (1 yes; 0 no)

Question aa will be asked if there is at least one uncorrelated random effect

aa) Save original labels to match with mean estimates for uncorrelated random effects in MTGSRUN (1 yes; 0 no)

If the option to write levels of fixed effects or uncorrelated random effects to unit 82 (MTGS82) is 1, summary statistics for each level will be written to the output log. With many levels of a fixed effect or uncorrelated random effect, answer no (0) to avoid a large output log.

On DOS or UNIX based systems, after gaining familiarity with the program, users may want to put the analysis information in a file for the program to read, which is easier than entering the data interactively. However, please enter the data interactively to become familiar with the

questions the first few times. If a mistake is made answering questions interactively, the program must be started from the beginning. To run the program using such an input file execute the program using the form:

mtgsprep.exe < input.fil

where mtgsprep.exe is the executable form of the MTGSPREP Fortran file and input.fil contains the same entries that would be entered interactively. If running the programs from a batch or script file it may be useful to use the command:

mtgsprep.exe < input.fil > output.fil

where output.fil is a file containing the prompts usually written to the terminal. An alternative approach is to change the file definition section for unit 5 in MTGSPREP to a physical file rather than keyboard input (i.e., change the value for IUN5 and add an open statement for that file). MTGSPREP produces the following files :

MTGS41 Information on model to be used in MTGSRUN. Information includes :

- number of traits, effects, animals, regression coefficients, equations, number of uncorrelated random effects, number of columns that contain uncorrelated random effects, number of fixed effects, column of animal ID in data set, whether the original labels for fixed effects and covariates were written to a file, and whether the original labels for uncorrelated random effects were written to a file
- name of each trait, number of covariates by trait; power of each covariate
- number of fixed effects by trait; number of levels for each fixed effect
- starting equation number for direct effects by trait
- presence of second animal (e.g., maternal genetic) effects by trait
- column of second animal ID in data set
- starting equation number for second animal effects by trait
- number of uncorrelated random effects by trait
- number of levels for each uncorrelated random group, column of uncorrelated random group in data set (if no. uncorrelated random effects > 0)
- starting equation number of uncorrelated random effects, uncorrelated random group number, and column positions from original data of each uncorrelated random effect by trait (if no. uncorrelated random effects > 0)

- ♦ name of each uncorrelated random effect by trait (if no. uncorrelated random effects > 0)

MTGS42 Recoded data for MTGSRUN in binary format. The file includes the coefficients of the "design" matrix **W** = [**X Z**], equation numbers, and the data deviated from the mean. To use the undeviated data search for the string 'DATA DEVIATION' in the program and comment out the appropriate write statement (a C in the first column of the line) and remove the comment for the line which writes the raw data.

MTGS43 Summary for each animal by record in binary format. The data include number of columns of W and number of observations written to MTGS42, a code representing the pattern of missing observations, and codes to determine if each trait is present or missing.

MTGS44 Blocking information used in MTGSRUN for related random effects coded in the same column as the animal or second animal ID for a trait. Data include animal ID, number of blocked equations, and equation numbers for effects to be blocked.

MTGS45 Original abels for covariates and fixed effects if requested for merging with posterior means in MTGSRUN.

MTGS46 Original labels for uncorrelated random effects if requested for merging with posterior means in MTGSRUN.

MTGS82 Program log that includes summary statistics and order and information about the mixed model equations.

Prior to running MTGSPREP, make sure that any output files to be saved from a previous MTGSPREP run are renamed or copied elsewhere. MTGSPREP will delete or overwrite output files written in earlier runs.

Estimating Variance Components or Solving MME (MTGSRUN)

The main function of MTGSRUN is to generate Gibbs samples for variance components and fixed and random effects under a flexible set of options. The program gives the user the option of using Gauss-Seidel iteration (**GSI**) to generate solutions to the MME using starting values for the variance components. The program can be used to obtain solutions only for the MME with appropriately chosen responses to the questions posed by MTGSRUN. If Gibbs sampling is requested (with or without GSI) the user has the option to specify the numbers of

rounds of Gibbs sampling ,burn-in, interval between Gibbs samples, and frequency of checkpointing. Checkpointing is the process of saving all necessary variables to files so that the program can be restarted at some intermediate step if the program is halted (accidentally or intentionally).

The first question to be asked by MTGSRUN is:

TYPE OF ANALYSIS:
1) New analysis
2) Continuation of Gauss-Seidel iteration using last solutions
3) Continuation of a previous analysis stopped prematurely
4) Continuation of a previous analysis stopped when completed

Option 1 is chosen to start any analysis, option 2 is used if further rounds of GSI are need to obtain a solution for the MME, option 3 is used if a Gibbs sampling analysis has been stopped for some reason and the analysis is to be continued, and option 4 allows additional rounds of Gibbs sampling if after preliminary analysis it is decided that more Gibbs samples are needed. The use of checkpointing allows all of these continuations to be used with little or no loss of accuracy. The continuation of a halted analysis should return the same analysis as run without interruption. The addition of rounds after completion may cause (minor) rounding differences in the posterior means of parameter estimates, but the sampled values should be identical to those run in one analysis. Any analysis must be initialized using option 1. Option 2 should be used only when no Gibbs sampling has been done following previous rounds of GSI. It is often useful to run analyses in off peak hours (e.g., nights and weekends). Option 3 gives users the option of running an analysis as computing resources permit, by using this option and checkpointing at appropriate intervals. Finally, the last option allows an analysis to be continued if it is decided that the initial Gibbs sampling chain length is insufficient.

Solving the MME using Gauss-Seidel iteration

The MME can be constructed and solved using a blocked Gauss-Seidel algorithm. The blocked algorithm updates all animal and second animal effects as well as blocked uncorrelated random effects simultaneously. An uncorrelated random effect is considered "blocked" if the code for that effect appeared in the same column as the animal ID or in the same column as the second animal effect for that trait (if one exists). All other effects (covariates, fixed effects, and

uncorrelated random effects not blocked) are updated one equation at a time. The program uses starting values of zeros for all fixed and random effects. Iterations are repeated until either a maximum number of iterations is reached or until the convergence criterion is met. The convergence criterion used is:

$$\frac{\sum_{i=1}^{n}\left(s_i^j - s_i^{j-1}\right)^2}{\sum_{i=1}^{n}\left(s_i^j\right)^2}$$

where n is the number of equations, and s_i^j is solution to equation i in round j of iteration.

Generating Gibbs samples

The main purpose of this set of programs is to generate values from a Gibbs sampler. The theory related to Gibbs sampling is given in Chapter 3, including discussion of the Bayesian variance component model, prior distributions, and derivation of the Gibbs sampling algorithm. A Bayesian variance component model is assumed, although in some situations it is possible to use flat priors for the variance components, and in that case the estimates should be similar to REML estimates. For a discussion of prior distributions including when it is safe to use flat priors for variance components see Chapter 3.

To run MTGSRUN

1. Compile and link MTGSRUN, MTGSSUB (general subroutines), and MTGSRSUB (subroutines used by MTGSRUN). GSPARAM.FOR must be available.
2. Run MTGSRUN. The program reads MTGS41, MTGS42, MTGS43, and MTGS44. The program needs MTGS45 and/or MTGS46 if merging of solutions and labels is requested.
   a) Enter option for analysis: 1, new analysis; 2, continuation of Gauss-Seidel iteration using last solutions; 3, continuation of a previous analysis stopped prematurely, or 4, continuation of a previous analysis stopped when completed.

   OPTION 1 QUESTIONS

   b) Enter description of the analysis terminated with a * in column 1 of the line following the last comment. The maximum number of lines of comments is defined in GSPARAM.FOR.
   c) Press enter to continue following user information. If an input file is used to for responses to questions be sure to insert a blank line in that file.

d) Input the mean of the prior distribution for additive and second animal (i.e., the G_o matrix) (co)variances. A matrix description of the (co)variance components is displayed which depends on the model chosen in MTGSPREP. The order of the genetic effects are animal effects first in order of traits followed by second animal effects, where appropriate. The row and column labels include an A or M for animal or second animal effect, respectively, followed by the trait number. For example, for a two trait model with second animal effects for only the first trait, the following information would be displayed:

	A1	A2	M1
A1:	1		
A2:	2	4	
M1:	3	5	6

Enter matrix position and values for priors.

For example, for σ^2_{a1}, an entry is :

1 100.D0 <return> [1 100. <return> or 1 100 <return> also work].

For σ_{A1A2};

2 -25.D0 <return>.

The matrix is initialized to zero. A prior is needed for each component to be estimated (0 is a valid estimate for covariances). Type -1 0.d0 <return> to show the position numbers again. Once all priors are entered, end the input by typing 0 0.d0 <return> [0 0 <return> also works].

e) The mean of the prior distribution will redisplay and verification is requested, enter: 0 if the values are not correct, 1 if the values are correct, or 2 to display the values again. If there are incorrect values only those values need to be re-entered, because the covariance matrix will not be initialized again.

f) Enter the shape parameter for this covariance matrix. The prior distribution for the covariance matrix is assumed to be an inverted Wishart distribution. The shape parameter corresponds to the degree of freedom parameter for the corresponding Wishart distribution. For a proper prior distribution the shape parameter must be at least 2 more than the order of the covariance matrix. To hold the covariance matrix constant use a value of -1 for the shape parameter. To use a "flat" prior enter a value of 0 for the shape parameter (see Chapter 3 to check if this is safe!).

g) Press enter to continue following user information (insert a blank line in input file).

h) The program will display any random effects coded in the same column in the original data set that appear in more than one trait. The user is given the opportunity to specify groups of traits which should have covariances restricted to zero. The trait and random effect names (as entered in MTGSPREP) are given. The user should answer with 0 if there are no covariances that need to be restricted and 1 if groups of traits are to be specified. An example of information provided by MTGSRUN follows:

TRAIT #	TRAIT NAME	RANDOM # W/IN TRT	RANDOM EFFECT NAME
1	Weaning Weight	1	Mat Perm Env
4	Yearling Weight	1	Mat Perm Env

Are there covariances among these random factors that should be ZERO?
0 = No (Covariances among all effects can be non-zero)
1 = Yes (Some covariances need to be restricted to zero)

Questions i and j repeated for each set of uncorrelated random effects to be split into groups.

i) Group numbers are requested for each combination of trait and random effect name. Groups must be consecutive starting with 1.

j) Group numbers will be redisplayed with trait and random effect name. Verification of group numbers is requested: 1=yes; 2=no

Questions k-m are repeated for each group of uncorrelated random effects (groups correspond to column used to code for the effect in the original data, or the number assigned to groups of uncorrelated randoms in a column as done in question i) .

k) Input mean of the prior distribution for uncorrelated random effect (co)variances for the group of uncorrelated random effects. A listing by group describing the trait and uncorrelated random effect with a code will be listed. Finally, a matrix description of the (co)variance components is displayed using those codes. An example of information provided by MTGSRUN follows:

CODE	TRAIT #	TRAIT NAME	RANDOM # W/IN TRT	RANDOM EFFECT NAME
I1	1	Weaning Weight	1	Mat Perm Env
I2	4	Yearling Weight	1	Mat Perm Env

Enter the expected values for the residual covariance matrix. I's correspond to the codes listed above.

	I1	I2
I1:	1	
I2:	2	3

Enter matrix position and values for priors.

l) The mean of the prior distribution will redisplay and verification is requested, enter: 0 if the values are not correct, 1 if the values are correct, or 2 to display the values again. If there are incorrect values only those values need to be re-entered, because the covariance matrix will not be initialized again.

m) Enter the shape parameter for this covariance matrix. For a proper prior the shape parameter must be at least 2 more than the order of the covariance matrix. To hold the covariance matrix constant use a value of -1 for the shape parameter. To use a "flat" prior enter a value of 0 for the shape parameter (see Chapter 3 to check if this is safe!).

If two or more traits are included in an analysis answer questions n-q.

n) Press enter to continue following user information (insert a blank line in input file).

o) The user is given the opportunity to specify groups of traits which should have residual covariances restricted to zero. The trait number and name are given. The user should answer with 0 if there are no covariances that need to be restricted and 1 if groups of traits are to be specified.

Questions p and q are asked if question o) is answered yes, that is if residual effects are to be split into groups.

p) Group numbers are requested for each trait. Groups must be consecutive starting with 1.

q) Group numbers will be redisplayed. Verification of group numbers is requested: 1=yes; 2=no

Questions r-t are repeated for each group of residual effects (groups correspond to traits in the original data, or the number assigned to groups of traits as done in question p) .

r) Input mean of the prior distribution for uncorrelated random effect (co)variances for the group of uncorrelated random effects. A listing by group describing the trait and uncorrelated random effect with a code will be listed. Finally, a matrix description of the (co)variance components is displayed using those codes. An example of information provided by MTGSRUN follows:

The residual effects coded in group 1 are represented in 3 trait(s) as follows:

CODE	TRAIT #	TRAIT NAME
R1	1	Pelvic Width
R2	3	Ovulation Rate
R3	4	Milk Production

Enter the expected values for the residual covariance matrix. R's correspond to the codes listed above

	R1	R2	R3
R1	1		
R2 :	2	4	
R3 :	3	5	6

Enter matrix position and values for priors.

s) The mean of the prior distribution will redisplay and verification is requested, enter: 0 if the values are not correct, 1 if the values are correct, or 2 to display the values again. If there are incorrect values only those values need to be re-entered, because the covariance matrix will not be initialized again.

t) Enter the shape parameter for this covariance matrix. For a proper prior the shape parameter must be at least 2 more than the order of the covariance matrix. To hold the covariance matrix constant use a value of-1for the shape parameter. To use a "flat" prior enter a value of 0 for the shape parameter (see Chapter 3 to check if this is safe!).

u) Enter the number of rounds of Gauss-Seidel iteration to be done before starting Gibbs sampling (Enter 0 for none).

v) Enter the convergence criterion for the Gauss-Seidel iteration. This value is the sum of squared changes insolutions divided by the sum of squared solutions. Use a relatively large value for warm up for Gibbs sampling (e.g., 1.D-3). Use a relatively small value for solutions to MMEs (e.g., 1.D-10).

w) Enter the length of the Gibbs sampling chain, including the burn-in period.

Questions x-ae are asked only if the number of rounds of Gibbs sampling is greater than zero.

x) Enter the length of the 'burn-in' period. This is the number of rounds of Gibbs sampling ignored before including data in means estimates or writing values to files.

y) Enter the frequency of writing data (variance components, contrasts, etc) to MTGS61 and MTGS62.

z) Enter the frequency of checkpointing. This is the frequency of writing critical information to allow a restart if stopped prematurely.

aa) Write all fixed and random effects at the same frequency as the variance component information? 0, no; 1, yes.

ab) Write a set of contrasts out at the same frequency as the variance component information? 0, no; 1, yes.

Questions ac-ae are asked if question ab is answered yes.

ac) Enter the number of contrasts to be monitored.

Questions ad-ae are asked for each contrast.

ad) Enter the number of elements in each contrast, e.g., 2.

Question ae is asked for each element in a contrast.

ae) Enter the equation numbers and coefficients in order for the contrast, e.g., 4 1. <return> 5 -1. <return>.

af) Enter two random number seeds. The first must be in the range 0 to 31328, and the second must be in the range 0 to 30081.

OPTION 2 QUESTIONS

The program will read information from the checkpoint files. All of the user respones from the previous analysis (entered with option 1) will be restored. Only the values for the variables affected by the following questions will be affected.

If there have been rounds of Gibbs sampling completed in the previous analysis after (or without) Gauss-Seidel iteration then the following information will be given and question b) will be asked:

Rounds of Gibbs sampling were completed after Gauss-Seidel iteration in the previous analysis. Do you want to continue Gauss-Seidel iteration from the current Gibbs sampling fixed and random solutions or restart the analysis?

** ALL PREVIOUS GIBBS SAMPLING DATA WILL BE LOST IF YOU CONTINUE OR RESTART!! **

b) Enter: 0 to stop the analysis
1 to restart the analysis
2 to proceed with Gauss-Seidel using final Gibbs sampling 'solutions'
3 to continue Gibbs sampling

If question b is not asked or if it is answered with option 2, the following questions c and d will be asked.

c) Enter the number of additional rounds of Gauss-Seidel iteration to be done before starting Gibbs sampling.
d) Enter the new convergence criterion for the Gauss-Seidel iteration. See question v under option 1 for information.

OPTION 3 QUESTIONS

There are no questions asked after the question for option number.

OPTION 4 QUESTIONS

If the previous analysis has been completed the program will ask question b. If the previous analysis was incomplete, it will reset the option to 3 and continue that anaylysis.

b) Enter the number of additional rounds of Gibbs sampling to be done.

Convergence

Measurement of convergence in Gibbs sampling is more difficult than with a likelihood based procedure. The main problem is that the convergence of a distribution must be evaluated, not the convergence to a single point as is the typical animal breeding problem (e.g., prediction of breeding values or estimation of variance components). There are several simple diagnostics which may be helpful. One alternative is to generate samples from multiple chains and compare the estimated means for parameters across chains or to estimate the parameters distribution for each chain and compare those estimates using a "fat tip pen test." The fat tip pen test is a subjective evaluation of the distributions to determine if the estimates are approximately equal,

i.e., does the line drawn by a fat tip pen cover both of the estimated distributions. If so, there is a reasonably good chance that the estimates are converged. The drawback of using multiple chains is that the burn-in period for each chain must be discarded and may correspond to significant increases in computing time for the same number of useful samples drawn from the Gibbs sampler.

Starting or Restarting MTGSRUN

On DOS or UNIX based systems, after gaining familiarity with the program, users may want to put the analysis information in a file for the program to read, which is easier than entering the data interactively. For restarts, users can copy MTGS60 to a restart input file. MTGS60 contains the original answers to the interactive questions except for the (co)variances, which are the mean estimates obtained at the end of the previous run. Please enter the data interactively to become familiar with the questions the first few times. If a mistake is made answering questions interactively, the program must be started from the beginning. To run the program using such an input file execute the program using the form:

```
mtgsrun.exe < input.fil
```

where mtgsrun.exe is the executable form of the MTGSRUN Fortran file and input.fil contains the same entries that would be entered interactively. If running the programs from a batch or script file it may be useful to use the command:

```
mtgsrun.exe < input.fil > output.fil
```

where output.fil is a file containing the prompts usually written to the terminal. An alternative approach is to change the file definition section for unit 5 in MTGSRUN to a physical file rather than keyboard input (i.e., change the value for IUN5 and add an open statement for that file).

CHAPTER TWO: Illustrations for MTGSAM

This chapter demonstrates models and analyses that can be run using MTDGSAM. The examples presented are based on the mouse data distributed with DFREML (Meyer, 1991). Data are available on diskette or by anonymous FTP. The data files are distributed with example single and multiple trait MTGSAM analyses. The format of the pedigree file, MOUSE.PED, includes three fields: animal, sire, and dam. The data file, MOUSE.DAT, contains ten fields of data - seven integer and three real. The integers correspond to: animal, sire, dam, generation, sex, litter size, and litter number. The three real fields represent: litter size (for use as a covariate), body weight, and feed intake.

The purpose of this section is to illustrate interactive sessions with the programs and the types of output generated as well as what to examine and expect from the output. All analyses demonstrated here were run on a Pentium class microcomputers with 64 MB of memory (although the analyses should run on systems with much less power and memory).

MTGSNRM

MOUSE.PED was the file used by MTGSNRM to produce the non-zero A^{-1} elements used in MTGSRUN. Two of the most important lines to note in the output file, MTGS81, are the number of pedigree lines read and the total number of different animals which is needed in MTGSPREP.

Results in MTGS81:

```
Started 10:49:39.96 on 04/06/1995

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
PROGRAM "MTGSNRM" - Calculate A-1 and recode animal for IDs for Gibbs sampling
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

 OPTION FOR CALCULATION OF A-1
     FOR ANIMAL   SIRE    DAM     TYPE ....    0
     FOR ANIMAL   SIRE    MGS     TYPE ...     1
 OPTION CHOSEN FOR THIS ANALYSIS                =              0  see note 1

 MAXIMUM ID                                     =          41615
 MINIMUM ID                                     =              1
 PEDIGREE FILE OPENED, IUN33                    = mouse.ped

 FILE FOR IDS AND INBREEDING COEFFICIENTS OPENED
   THIS FILE WILL CONTAIN ANIMAL, SIRE, AND DAM
   RECODED AND ORIGINAL IDS FOLLOWED BY THE
   INBREEDING COEFFICIENT FOR EACH

 NO. INTEGER FIELDS PER RECORD IN IUN33         =              4
 ANIMAL ID IN POSITION ......                                  1
 SIRE ID IN POSITION ........                                  2
 DAM (MGS) ID IN POSITION ...                                  3
 NO. OF GENETIC GROUPS FOR CALCULATION OF W     =              0

 The current time is:  10:49:40.29

 NO. OF PEDIGREES READ                          =            309  see note 2
 NO. OF DIFFERENT ANIMALS                       =            329  see note 3
 INCLUDES NO. OF GENETIC GROUPS                 =              0

 END OF FIRST PASS
 The current time is:  10:49:40.45

 END OF SORT
 The current time is:  10:49:40.45

 FIRST 10 REORDERED IDs          1         215                    see note 4
 FIRST 10 REORDERED IDs          2         403
 FIRST 10 REORDERED IDs          3         615
 FIRST 10 REORDERED IDs          4         701
```

Note 1: The answers highlighted in gray were answers to the interactive question asked by MTGSAM. Check to make sure that they are correct

Note 2: Does this agree with your data? This number should equal number of data lines in pedigree file. Animals can be repeated in data file.

Note 3: This is the number of animals plus the number of base animals. Make sure that the number of base animals is at least 0. The number of base animals is the number of different animals minus the number of pedigrees read.

Note 4: Reordered animal identification numbers with original animal identification. These animal IDs should be reasonable given the data set.

```
FIRST 10 REORDERED IDs          5         814
FIRST 10 REORDERED IDs          6         904
FIRST 10 REORDERED IDs          7        1314
FIRST 10 REORDERED IDs          8        1602
FIRST 10 REORDERED IDs          9        1701
FIRST 10 REORDERED IDs         10        1813

ID VECTOR WRITTEN IN ORDER TO IUN21
The current time is:  10:49:40.51

SIRE AND DAM IN PEDIGREE REORDERED IN IVECS AND IVECD
The current time is:  10:49:40.56

CALCULATION OF A-1 FROM ANIMAL  SIRE  DAM  (IOPT =   0)

NON-ZERO HS ELEMENTS FOR NRM INVERSE            =           1241
LOG DETERMINANT OF NRM                          =  -210.71674289
NUMBER OF INBRED ANIMALS                        =              0
 ... WITH AVERAGE INBREEDING COEF               =      .00000000
TOTAL NO. OF ANIMALS INCLUDING BASE
      AND GENETIC GROUPS                        =            329  see note 3

The current time is:  10:49:40.73
The elapsed time was: 00:00:00.44
```

Variance Component Estimation

Single Trait Model

The data for mouse body weight were analyzed with a model including additive (direct) genetic effect, correlated second animal genetic effect and one uncorrelated random effect. The data include 284 observations for body weight in mice. Additive direct genetic effect of animal, maternal genetic effect of second animal (the dam) and a maternal permanent environmental effect are in the model. Three fixed effects were: generation, sex and litter size.

MTGSPREP

For this example, the option to write levels of information to MTGS81 for all fixed effects was enabled and for the uncorrelated random effects was disabled. The complete list of answers to the interactive questions follow.

mouse.dat	name of data file
Mouse data from Karin Meyer	
Single trait analysis of body weight	
*	end of comments
7	number of integers on each line of data file

3	number of reals on each line of data file
1	number of traits in analysis
body weight	name of trait 1
2	position of trait in vector of real values
0.0	value of missing observation for trait 1
0	number of covariates
3	number of fixed factors
generation	name for fixed factor 1
4	position of fixed factor 1 in vector of integers
1	write summary of fixed factor 1 levels to log file (MTGS82)
sex	name for fixed factor 2
5	position of fixed factor 2 in vector of integers
1	write summary of fixed factor 2 levels to log file (MTGS82)
litter size	name for fixed factor 3
6	position of fixed factor 3 in vector of integers
1	write summary of fixed factor 3 levels to log file (MTGS82)
1	position of animal effect in vector of integers
329	num. of animals in relationship matrix (from MTGSNRM)
1	include second animal effect
maternal genetic	name of second animal effect
3	position of second animal effect in vector of integers
1	number of uncorrelated random factors
maternal perman env	name of uncorrelated random factor
3	position of uncorrelated random factor in vector of integers
0	do not write summary of uncorrelated random factor to log
1	write summary of fixed factor 1 levels to log file (MTGS82)
1	write labels for uncorrelated random factors to MTGS46

Results in MTGS82:

```
Started 13:30:31.47 on 04/06/1995
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
                PROGRAM "MTGSPREP"  - Setup MME for Gibbs sampling
                          Last revised ALPHA VERSION
+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
 Data set description :
    Mouse data from Karin Meyer
    Single trait analysis

 No. of data lines in Unit 31           =          284          see note 5
 No. of integer variables per record    =            7
 No. of real variables per record       =            3
 No. of traits                          =            1
 No. of valid records                   =          284
 No. of animals with valid records      =          284
 No. of animals in A-1                  =          329
 Order of MME                           =          712
```

Note 5: Does this correspond to the data?

```
 Results for trait  1  - Body Weight           (position  2 )      see note 6

 No. of records =    284  (missing value:      .0000   No. missing =     0 )
 Trait   Mean       SD        CV       Min       Max     Std Min    Std Max
   1    24.0687  3.30236     13.72   14.600    34.500     -2.87       3.16

 No. of covariates =      0

 No. of fixed effects =      3
  1:     3 levels for  generation                    (MME rows:     1 -     3)
    Level     Value        No.        %       Mean
       1          1         93      32.75    23.724                 see note 6
       2          2         84      29.58    23.063
       3          3        107      37.68    25.158
  2:     2 levels for  sex                           (MME rows:     4 -     5)
    Level     Value        No.        %       Mean
       1          1        150      52.82    22.656                 see note 6
       2          2        134      47.18    25.650
  3:     7 levels for  litter size                   (MME rows:     6 -    12)
    Level     Value        No.        %       Mean
       1          1         11       3.87    26.609                 see note 6
       2          2         41      14.44    23.722
       3          3         25       8.80    24.864
       4          4         36      12.68    24.028
       5          5         96      33.80    24.265
       6          6         45      15.85    24.333
       7          7         30      10.56    21.973

 No. of animals in A-1 =   329                       (MME rows:    13 -   341)

 No. of 2nd animal effects =     1                                 see note 7
  1:   329 levels for  maternal genetic              (MME rows:   342 -   670)

 No. of uncorrelated random effects =     1                       see note 8
  1:    42 levels for  maternal perman env          (MME rows:   671 -   712)

-------------------------------------------------------------------------------
                  Summary of data and mixed model equations
-------------------------------------------------------------------------------
Trait    1 - Body Weight          No. of records =    284  (No. missing = 0 )

 Trait   Mean       SD        CV       Min       Max     Std Min    Std Max
   1    24.0687  3.30236     13.72   14.600    34.500     -2.87       3.16

 Order of MME =     712
```

Note 6: Are these characteristics of your data reasonable?
Note 7: An equation is created for the second animal effect for all animals
Note 8: An equation is created for each level of an uncorrelated random effect

see note 9

```
 Number of fixed effects =      3
 Trait      No.      Name               Position    Levels         Rows
    1        1     generation              4          3          1 -      3
    1        2     sex                     5          2          4 -      5
    1        3     litter size             6          7          6 -     12

 Number of animal effects (# traits) =      1
 Trait      No.      Name               Position    Levels         Rows
    1        1     Animal w/ full A-1      1         329        13 -    341


 Number of second animal effects =      1
 Trait      No.      Name               Position    Levels         Rows
    1        1     maternal genetic        3         329        342 -    670


 Number of uncorrelated random effects =      1
 Trait      No.      Name               Position    Levels         Rows
    1        1     maternal perman env     3          42        671 -    712
-------------------------------------------------------------------------------

 Files written:
   MTGS41 (ascii): Model information
   MTGS42 (binary): Recoded W=X:Z elements
   MTGS43 (binary): W summary for each animal
   MTGS44 (binary): Blocking information by animal
   MTGS45 (ascii): labels for covariates and fixed effects
   MTGS46 (ascii): labels for uncorrelated random effects
 The elapsed time was: 00:00:00.54
```

Note 9: Check number of levels and positions of fields in integer vector for possible input errors and order of MME

MTGSRUN

Answers to the interactive questions asked by MTGSRUN:

1	type of run - new analysis
Mouse data from Karin Meyer	
Single trait analysis of body weight	
*	
	blank line needed for press enter prompt
1 4.0	genetic (co)variance means
2 0.5	genetic (co)variance means
3 1.5	genetic (co)variance means
0 .00	done entering values
1	genetic priors are correct
9	genetic (co)variance priors shape parameter
	blank line needed for press enter prompt
1 1.5	independent random (co)variance means
0 .00	done entering values

```
1              ind Rand priors are correct
9              ind Rand (co)variance priors shape parameter
1  2.0         residual (co)variance priors
0  .00         done entering values
1              residual priors are correct
9              residual (co)variance priors shape parameter
100            rounds of Gauss-Seidel iteration
1.E-04         convergence criteria for Gauss-Seidel
200            rounds of Gibbs sampling (includes burn-in)
100            rounds of burn-in before Gibbs sampling
10             frequency of writing Gibbs samples
50             frequency of check-pointing
0              write out all solutions? (Y=1,N=0)
0              write out user specified contrasts
29193  21661   seeds for random number generator
```

Note that the number of rounds of Gibbs sampling is **VERY SMALL** and is used here just for demonstration purposes. The burn-in and frequency of writing samples are also chosen only for demonstration purposes.

Results in MTGS83:

```
--------------------------------------------------------------------------------
                "MTGSRUN" - Multiple trait Gibbs sampling program
                          Last revised ALPHA VERSION
--------------------------------------------------------------------------------

 Started 08:01:39.81 on 04/14/1995

 Mouse data from Karin Meyer
 Single trait analysis of body weight

 This is a new analysis - not a continuation

 The prior distribution of genetic variances and covariances was an
 inverted Wishart distribution with shape parameter:   9 and with
 expected value:

           A1           M1
 A1 :   4.0000
 M1 :   .50000       1.5000

 The prior distribution of variances and covariances for the random
 factors coded in column   3 and represented in   1 trait(s) was an
 inverted Wishart distribution with shape parameter:   9 and
 with expected value:
```

```
          I1
I1 :    1.5000

The I's correspond to the codes below

                                    RANDOM #
 CODE  TRAIT #  TRAIT NAME         W/IN TRT  RANDOM EFFECT NAME
  I1      1     Body Weight            1     maternal perman env

The prior distribution of variances and covariances for the residual
effects was an inverted Wishart distribution with shape parameter:   9
and with expected value:

          R1
R1 :    2.0000

The R's correspond to the codes below

 CODE  TRAIT #  TRAIT NAME
  R1      1     Body Weight

There was a maximum of     100 rounds of Gauss-Seidel iteration done
before starting Gibbs sampling.

A value of .10000D-03 was used to determine convergence of
Gauss-Seidel iteration

There were    100 rounds burn-in run before using the Gibbs
sampling done before using the results

There were    200 total rounds of Gibbs sampling including the burn-in period

Results were written out every    10 rounds.

Checkpointing information was written out every     50 rounds.

The two random number generator seeds used were: 29193 and 21661

Gauss-Seidel iteration converged in round     13
with a convergence criteria of  .849214D-04

For the genetic and independent random effect and residual
(co)variance components several estimates are provided.  The first
column contains the posterior mean of the expected value of the
component.  The second is the posterior mean of the observed values.
The final column contains the posterior mean of the observed values
for heritability (genetic effects) or fraction of phenotypic variance
(independent randoms and residual) on the diagonal and correlations
below the diagonal.  For phenotypic (co)variance components only the
posterior means of the observed components and the correlations are
given.

If a (co)variance has been held constant, only the values and the
means of the observed "heritabilities" and correlations are given.
```

GENETIC (CO)VARIANCE COMPONENT ESTIMATES

A correspond to direct effects, M to second animal (e.g., maternal) effects
For example, M3 is second animal genetic effect for trait 3

```
           A1    M1
  A1 :      1
  M1 :      2     3
```

VC	POST MEAN OF EXP VC	POST MEAN OF OBS VC	POST MEAN OF OBS CORR
1	4.3072971277831	4.2662797793787	.382937737008567800
2	1.7888933877744	1.7707084304743	.630660694294524500
3	1.8796056518771	1.8795230673661	.169380671693390900

INDEPENDENT RANDOM EFFECT (CO)VARIANCE COMPONENT ESTIMATES

The random factors coded in column 3 are represented in 1 trait(s) as follows:

CODE	TRAIT #	TRAIT NAME	RANDOM # W/IN TRT	RANDOM EFFECT NAME
I1	1	Body Weight	1	maternal perman env

I's correspond to the codes listed above

```
           I1
  I1 :      1
```

VC	POST MEAN OF EXP VC	POST MEAN OF OBS VC	POST MEAN OF OBS CORR
1	1.1976811419801	1.2080489869830	.110473240469843800

RESIDUAL (CO)VARIANCE COMPONENT ESTIMATES

The residual effects are represented in 1 trait(s) as follows:

CODE	TRAIT #	TRAIT NAME
R1	1	Body Weight

R's correspond to the codes listed above

```
           R1
  R1 :      1
```

VC	POST MEAN OF EXP VC	POST MEAN OF OBS VC	POST MEAN OF OBS CORR
1	1.9305686899210	1.9370950692461	.179021126890384100

PHENOTYPIC (CO)VARIANCE COMPONENT ESTIMATES

CODE	TRAIT #	TRAIT NAME
P1	1	Body Weight

```
 P's correspond to the traits listed above

            P1
  P1 :       1

 VC       POST MEAN OF OBSERVED VC       POST MEAN OF OBSERVED CORR
 ==     ==========================     ==========================
  1           11.06165533344822            1.00000000000000000000
```

Multiple Trait Model

The data for mouse body weight and feed intake were analyzed with a multiple trait animal model. The model for body weight included additive (direct) genetic effect and one uncorrelated random effect. The data include 284 observations for body weight. Additive direct genetic effect of animal and a litter effect are in the model. One covariate and one fixed effect were fit. Litter size was included as a covariate and generation was considered a fixed effect. The model for feed intake also included direct genetic effect and one uncorrelated random effect. There were two fixed effects included: litter size and generation.

Note that litter size is included as a covariate for one trait and a fixed effect for the other. The use of factor as a covariate and fixed effect is possible because there are two fields set for the same effect - one in the vector of integers and one in the vector of real values. If the same effect is not to be used for two traits in the same analysis the same field can be used as a covariate in one analysis and a fixed effect in a second analysis by placing the field in a location where it can be included in the real values for the first analysis and as an integer in the second analysis.

MTGSPREP

For this example, the option to write summary information to MTGS81 for levels of fixed factors was enabled for all fixed effects. The option was enabled for the uncorrelated random effect for body weight and disabled for feed intake. The complete list of answers to the interactive questions follow.

```
mouse.dat               name of data file
Mouse data from Karin Meyer
Multiple trait analysis of
Body Weight and Feed Intake
*                       end of comments
7                       number of integers on each line of the data file
```

3	number of reals on each line of the data file	
2	number of traits in the analysis	
Body Weight	name of trait 1	Trait 1
2	position of trait 1 in vector of real values	
0.	value of missing observation for trait 1	
1	number of covariates for trait 1	
Litter Size	name of covariate 1	
1	position of covariate in vector of real values	
1	maximum power of covariate	
1	number of fixed effects for trait 1	
Generation	name of fixed effect 1	
4	position of fixed effect 1 in vector of integers	
1	write summary of fixed effect 1 levels to log file (MTGS82)	
1	position of animal effect in vector of integers	
329	number of animals in relationship matrix (from MTGSNRM)	
0	include second animal effect for trait 1: 1=yes; 0=no	
1	number of uncorrelated random effects for trait 1	
Litter	name of uncorrelated random effect	
7	position of uncorrelated random effect in vector of integers	
1	write summary of uncorrelated random effect to log file (MTGS82)	
Feed Intake	name of trait 2	Trait 2
3	position of trait 2 in vector of reals	
0.	value of missing observation for trait 2	
0	number of covariates for trait 2	
2	number of fixed effects for trait 2	
Litter Size	name of fixed effect 1	
6	position of fixed effect 1 in vector of integers	
1	write summary of fixed effect 1 levels to log file (MTGS82)	
Sex	name of fixed effect 2	
5	position of fixed effect 1 in vector of integers	
1	write summary of fixed effect 2 levels to log file (MTGS82)	
0	include second animal effect for trait 1: 1=yes; 0=no	
1	number of uncorrelated random effects for trait 1	
Litter	name of uncorrelated random effect	
7	position of uncorrelated random effect in vector of integers	
0	do not write sum. of uncorr. random effect to log file (MTGS82)	
1	write labels for covariates and fixed effects to MTGS45	
1	write labels for uncorrelated random effects to MTGS46	

Results in MTGS82:

```
 Started 11:36:13.52 on 04/14/1995

++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
                PROGRAM "MTGSPREP"  - Setup MME for Gibbs sampling
                          Last revised ALPHA VERSION
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

 Data set description :
    Mouse data from Karin Meyer
    Multiple trait analysis of
    Body Weight and Feed Intake

 No. of data lines in Unit 31            =            284
 No. of integer variables per record     =              7
 No. of real variables per record        =              3
 No. of traits                           =              2
 No. of valid records                    =            568
 No. of animals with valid records       =            284
 No. of animals in A-1                   =            329
 Order of MME                            =            755

--------------------------------------------------------------------------------

 Results for trait  1  - Body Weight           (position  2 )

 No. of records =    284  (missing value:      .0000   No. missing =     0 )
 Trait   Mean       SD        CV        Min         Max     Std Min    Std Max
   1    24.0687  3.30236    13.72      14.600      34.500    -2.87       3.16


 No. of covariates =      1
  1:     1 regression coefficients for Litter Size  (MME rows:     1 -      1)

 Statistics for covariates:
 Cov.    Mean       SD        CV        Min         Max     Std Min    Std Max
  1    4.47887  1.64829      36.80     1.0000     7.0000     -2.11       1.53

 No. of fixed effects =      1
  1:     3 levels for  Generation                    (MME rows:     2 -     4)
    Level     Value        No.        %       Mean
       1          1         93      32.75    23.724
       2          2         84      29.58    23.063
       3          3        107      37.68    25.158

 No. of animals in A-1 =   329                      (MME rows:    14 -   342)

 No. of 2nd animal effects =     0

 No. of uncorrelated random effects =     1
  1:    42 levels for  Litter                       (MME rows:   672 -   713)
    Level     Value        No.        %       Mean
      1         1         8      2.82    23.800
      2         2         7      2.46    23.014
      3         3         5      1.76    22.880
      4         4         7      2.46    24.129
      5         5         8      2.82    21.687
      6         6         8      2.82    18.300
```

7	7	7	2.46	25.471
8	8	7	2.46	27.186
9	9	7	2.46	23.514
10	10	8	2.82	28.375
11	11	7	2.46	22.943
12	12	6	2.11	23.467
13	13	8	2.82	23.750
14	14	8	2.82	24.212
15	15	4	1.41	25.400
16	16	8	2.82	19.113
17	17	6	2.11	27.317
18	18	8	2.82	23.850
19	19	6	2.11	24.750
20	20	6	2.11	25.067
21	21	4	1.41	21.925
22	22	6	2.11	21.033
23	23	5	1.76	24.180
24	24	8	2.82	23.537
25	25	7	2.46	22.386
26	26	8	2.82	19.462
27	27	7	2.46	24.843
28	28	7	2.46	25.429
29	29	8	2.82	25.175
30	30	8	2.82	22.950
31	31	7	2.46	25.457
32	32	8	2.82	23.450
33	33	6	2.11	23.983
34	34	2	.70	32.650
35	35	7	2.46	26.757
36	36	5	1.76	25.160
37	37	6	2.11	25.167
38	38	7	2.46	26.414
39	39	7	2.46	24.443
40	40	6	2.11	24.317
41	41	8	2.82	26.688
42	42	8	2.82	25.063

Results for trait 2 - Feed Intake (position 3)

No. of records = 284 (missing value: .0000 No. missing = 0)

Trait	Mean	SD	CV	Min	Max	Std Min	Std Max
2	64.2556	5.93258	9.23	46.900	82.100	-2.93	3.01

No. of covariates = 0

No. of fixed effects = 2

1: 7 levels for Litter Size (MME rows: 5 - 11)

Level	Value	No.	%	Mean
1	1	11	3.87	58.355
2	2	41	14.44	61.578
3	3	25	8.80	65.020
4	4	36	12.68	61.531
5	5	96	33.80	65.286
6	6	45	15.85	66.242
7	7	30	10.56	66.433

2: 2 levels for Sex name (MME rows: 12 - 13)

Level	Value	No.	%	Mean
1	1	150	52.82	61.392
2	2	134	47.18	67.461

```
 No. of animals in A-1 =   329                        (MME rows:   343 -   671)

 No. of 2nd animal effects =     0

 No. of uncorrelated random effects =     1
  1:    42 levels for  Litter                         (MME rows:   714 -   755)

--------------------------------------------------------------------------------
            Summary of data and mixed model equations
--------------------------------------------------------------------------------

 Trait    1 - Body Weight      No. of records =    284  (No. missing =     0 )
 Trait    2 - Feed Intake      No. of records =    284  (No. missing =     0 )


 Trait   Mean       SD        CV        Min         Max     Std Min    Std Max
   1    24.0687   3.30236    13.72     14.600      34.500    -2.87       3.16
   2    64.2556   5.93258     9.23     46.900      82.100    -2.93       3.01

 Order of MME =     755

 Number of covariates =      1
 Trait      No.      Name               Position    Coeff.          Rows
    1        1     Litter Size             1          1          1 -      1

 Number of fixed effects =      3
 Trait      No.      Name               Position    Levels          Rows
    1        1     Generation              4          3          2 -      4
    2        1     Litter Size             6          7          5 -     11
    2        2     Sex            name     5          2         12 -     13

 Number of animal effects (# traits) =      2
 Trait      No.      Name               Position    Levels          Rows
    1        1     Animal w/ full A-1      1         329         14 -    342
    2        1     Animal w/ full A-1      1         329        343 -    671

 Number of uncorrelated random effects =      2
 Trait      No.      Name               Position    Levels          Rows
    1        1     Litter                  7          42        672 -    713
    2        1     Litter                  7          42        714 -    755

--------------------------------------------------------------------------------

 Files written:
   MTGS41 (ascii): Model information
   MTGS42 (binary): Recoded W=X:Z elements
   MTGS43 (binary): W summary for each animal
   MTGS44 (binary): Blocking information by animal
   MTGS45 (ascii): labels for covariates and fixed effects
   MTGS46 (ascii): labels for uncorrelated random effects

 The elapsed time was: 00:00:00.66
```

MTGSRUN

Answers to the interactive questions asked by MTGSRUN.

```
1                type of run - new analysis
Mouse data from Karin Meyer
Multiple trait analysis of
Body Weight and Feed Intake
*
                 blank line needed for press enter prompt
1       7.87     animal effect prior value (σ²A1)
2       2.50     animal effect prior value (σA1,A2)
3       9.99     animal effect prior value (σ²A2)
0       0        end of genetic (co)variance input
1         values are correct: 0=no, 1=yes, 2=redisplay
9                genetic (co)variance priors shape parameter
                 blank line needed for press enter prompt
0                covariances among groups of uncorr. randoms to restrict to zero? (0=no, 1=yes)
1       1.38     uncorrelated effect starting value (σ²C1)
2      -1.58     uncorrelated effect starting value (σC1,C2)
3       2.98     uncorrelated effect starting value (σ²C2)
0       0        end of uncorrelated random (co)variance input
1         values are correct
9                uncorrelated random (co)variance priors shape parameter
                 blank line needed for press enter prompt
0                covariances among groups of uncorr. randoms to restrict to zero? (0=no, 1=yes)
1       2.66     residual effect starting value (σ²R1)
2       2.61     residual effect starting value (σR1,R2)
3      11.13     residual effect starting value (σ²R2)
0      0 end of residual (co)variance input
1         values are correct
9                residual (co)variance priors shape parameter
200              rounds of Gauss-Seidel iteration
1.d-5            convergence criteria for Gauss-Seidel
500              rounds of Gibbs sampling (including burn-in)
100              rounds of burn-in before writing samples
20               frequency of writing Gibbs samples
10               frequency of checkpointing
0                write out all solutions with samples (0=no, 1=yes)
0                write out specified contrasts with samples (0=no, 1=yes)
8939 20902       seeds for random number generator
```

Note that similar to the single trait example, the number of rounds of Gibbs sampling is **VERY SMALL** and is used here just for demonstration purposes. The burn-in and frequency of writing samples are also chosen only for demonstration purposes.

Results in MTGS83:

```
--------------------------------------------------------------------------------
               "MTGSRUN" - Multiple trait Gibbs sampling program
                         Last revised ALPHA VERSION
--------------------------------------------------------------------------------

 Started 14:10:08.26 on 04/14/1995

 Mouse data from Karin Meyer
 Multiple trait analysis of
 Body weight and feed intake

 This is a new analysis - not a continuation

 The prior distribution of genetic variances and covariances was an
 inverted Wishart distribution with shape parameter:   9 and with
 expected value:

           A1           A2
A1 :   7.8700
A2 :   2.5000       9.9900

 The prior distribution of variances and covariances for the random
 factors coded in column   7 and represented in   2 trait(s) was an
 inverted Wishart distribution with shape parameter:   9 and
 with expected value:

           I1           I2
I1 :    1.3800
I2 :   -1.5800       2.9800

 The I's correspond to the codes below

                                      RANDOM #
  CODE  TRAIT #  TRAIT NAME          W/IN TRT  RANDOM EFFECT NAME
   I1      1     Body Weight             1     Litter
   I2      2     Feed Intake             1     Litter

 The prior distribution of variances and covariances for the residual
 effects was an inverted Wishart distribution with shape parameter:   9
 and with expected value:

           R1           R2
R1 :    2.6600
R2 :    2.6100       11.310

 The R's correspond to the codes below
```

```
 CODE  TRAIT #  TRAIT NAME
  R1      1     Body Weight
  R2      2     Feed Intake

There was a maximum of     200 rounds of Gauss-Seidel iteration done
before starting Gibbs sampling.

A value of  .100000D-04 was used to determine convergence of
Gauss-Seidel iteration

There were    100 rounds burn-in run before using the Gibbs
sampling done before using the results

There were    500 total rounds of Gibbs sampling including
the burn-in period

Results were written out every     20 rounds.

Checkpointing information was written out every     10 rounds.

The two random number generator seeds used were:  8939 and 20902

Gauss-Seidel iteration converged in round    125
with a convergence criteria of  .986327D-05

For the genetic and independent random effect and residual
(co)variance components several estimates are provided.  The first
column contains the posterior mean of the expected value of the
component.  The second is the posterior mean of the observed values.
The final column contains the posterior mean of the observed values
for heritability (genetic effects) or fraction of phenotypic variance
(independent randoms and residual) on the diagonal and correlations
below the diagonal.  For phenotypic (co)variance components only the
posterior means of the observed components and the correlations are
given.

If a (co)variance has been held constant, only the values and the
means of the observed "heritabilities" and correlations are given.


GENETIC (CO)VARIANCE COMPONENT ESTIMATES
A correspond to direct effects, M to second animal (e.g., maternal) effects
For example, M3 is second animal genetic effect for trait 3
           A1    A2
  A1 :      1
  A2 :      2     3

VC    POST MEAN OF EXP VC     POST MEAN OF OBS VC    POST MEAN OF OBS CORR
==   ====================   ====================   ====================
 1     7.6128712242734         7.6244012978318         .625885316788609700
 2     2.6353199371560         2.6135917088905         .317260115589680000
 3     8.6969256552332         8.6940402984235         .361645393507590000


INDEPENDENT RANDOM EFFECT (CO)VARIANCE COMPONENT ESTIMATES
The random factors coded in column   7 are represented in   2
trait(s) as follows:
```

CODE	TRAIT #	TRAIT NAME	RANDOM # W/IN TRT	RANDOM EFFECT NAME
I1	1	Body Weight	1	Litter
I2	2	Feed Intake	1	Litter

I's correspond to the codes listed above

	I1	I2
I1 :	1	
I2 :	2	3

VC	POST MEAN OF EXP VC	POST MEAN OF OBS VC	POST MEAN OF OBS CORR
1	1.5107549042936	1.5015300872335	.123695539121797300
2	-1.6584116514526	-1.6350729678939	-.752155531137562700
3	3.2795626422993	3.2759175761303	.136756123538846000

RESIDUAL (CO)VARIANCE COMPONENT ESTIMATES
The residual effects are represented in 2 trait(s) as follows:

CODE	TRAIT #	TRAIT NAME
R1	1	Body Weight
R2	2	Feed Intake

R's correspond to the codes listed above

	R1	R2
R1 :	1	
R2 :	2	3

VC	POST MEAN OF EXP VC	POST MEAN OF OBS VC	POST MEAN OF OBS CORR
1	2.9159545081281	2.9374446952519	.250419144089593200
2	2.5894656720673	2.5956383897162	.440715557884698400
3	11.932969999529	11.919454743307	.501598482953563200

PHENOTYPIC (CO)VARIANCE COMPONENT ESTIMATES

CODE	TRAIT #	TRAIT NAME
P1	1	Body Weight
P2	2	Feed Intake

P's correspond to the traits listed above

	P1	P2
P1 :	1	
P2 :	2	3

VC	POST MEAN OF OBSERVED VC	POST MEAN OF OBSERVED CORR
1	12.06337608031729	1.00000000000000000000
2	3.574157130712805	.21073590131847460000
3	23.88941261786115	1.00000000000000000000

Contrasts

The multiple trait example will be used to demonstrate the use of contrasts. Unlike the MTDFREML program, when using the MTGSAM program the contrasts must be specified at the beginning of an analysis. This is because the contrast value is written out as the Gibbs samples are written rather than obtained as a solution to the mixed model equations using the final variance component estimates. As a result, the contrasts account for uncertainty of other parameters in the model when the distribution is estimated. It is important that the contrast be estimable. There is no tool to determine expected values of solutions; the MTDFREML programs may be used to evaluate expectations and estimability.

Only the changes in the input data will be given. The contast information is written to MTGS83, and only the information additional to that previous given will be presented. The following lines replace the line on page 38, 42 with the label "write out specified contrasts with samples."

```
1          write out specicified contrasts with samples: 0=no; 1=yes
4          number of contrasts
1          number of elements in the contrast 1
1    1.    equation number of solution for contrast and coefficient for element 1
2          number of elements in the contrast 2
4    1.    equation number of solution for contrast and coefficient for element 1
2   -1.    equation number of solution for contrast and coefficient for element 2
2          number of elements in the contrast 3
13   1.    equation number of solution for contrast and coefficient for that solution
12  -1.    equation number of solution for contrast and coefficient for that solution
1          number of elements in the contrast 4
14   1.    equation number of solution for contrast and coefficient for that solution
```

CHAPTER THREE: Theoretical Considerations for MTGSAM

Mixed Models

In matrix notation the general mixed model for an observation vector, **y**, is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \text{ where}$$

$\boldsymbol{\beta}$ = vector of fixed effects associated with records in **y** by **X**, and

u = vector of random effects associated with records in **y** by **Z**,

$$E\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{X}\boldsymbol{\beta}\\ \mathbf{0}\\ \mathbf{0}\end{bmatrix}, \text{ and } \mathrm{Var}\begin{bmatrix}\mathbf{y}\\ \mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'+\mathbf{R} & \mathbf{Z}\boldsymbol{\Sigma} & \mathbf{R}\\ \boldsymbol{\Sigma}\mathbf{Z}' & \boldsymbol{\Sigma} & \mathbf{0}\\ \mathbf{R} & \mathbf{0} & \mathbf{R}\end{bmatrix},$$

$$\boldsymbol{\Sigma} = \begin{bmatrix}(\mathbf{G}\otimes\mathbf{A}) & \mathbf{0}\\ \mathbf{0} & \bigoplus_{i=1}^{\gamma}\left[\mathbf{D}_i \otimes \mathbf{I}_{r_i}\right]\end{bmatrix}, \text{and}$$

$$\mathbf{R} = \bigoplus_{j=1}^{n}\mathbf{R}_j^*,$$

where i corresponds to the group of uncorrelated random effects which are common across traits (e.g., maternal permanent environmental effects), $\mathbf{D}_i$ describes the (co)variances among those random effects across traits for an animal, $\mathbf{R}_j^*$ is the matrix of (co)variances of residuals for the traits measured on an animal j, r_i is the number of levels of random group i, n is the number of animals, and $\otimes$ and $\oplus$ correspond to the direct product and direct sum operators, respectively (see Searle (1982) for a description of these operators). Note that for an animal with all traits measured $\mathbf{R}_i^* = \bigoplus_{j=1}^{\rho}\mathbf{R}_j$, where $\mathbf{R}_j$ is the $t_j \times t_j$ covariance matrix among residuals for block j of the residual effects and ρ is the number of blocks of residual effects. The blocks correspond to uncorrelated random effects for different traits that are coded in the same column in the original data (e.g., permanent environmental effects in multiple traits). The blocks for uncorrelated random and residual effects can be divided by the user when GSPREP is run. The division of the blocks correspond to groups of traits that are not observed on the same animal, e.g., sex limited traits such as milk production and scrotal circumference.

Define $\boldsymbol{\Sigma}_0 = \mathbf{G} \otimes \mathbf{A}$ and $\boldsymbol{\Sigma}_i = \mathbf{D}_i \otimes \mathbf{I}_{r_i}$, then $\boldsymbol{\Sigma} = \bigoplus_{i=0}^{\gamma} \boldsymbol{\Sigma}_i$, where γ is the number of groups of uncorrelated random effects. In addition, define $\mathbf{u}' = \left[\mathbf{u}_0' \, \mathbf{u}_1' \cdots \mathbf{u}_\gamma'\right]$, where $\mathbf{u}_0$ corresponds to the genetic effects, and $\mathbf{u}_i$ to the uncorrelated random effects in block i, for i > 0.

In many animal breeding applications for a single trait analysis, $\mathbf{u}$ is a vector of breeding values with $V(\mathbf{u}) = \mathbf{G} = \mathbf{A}\sigma_g^2$, where $\mathbf{A}$ is the numerator relationship matrix and σ_g^2 is the additive genetic variance (variance of breeding values) and $\mathbf{R} = \mathbf{I}\sigma_e^2$.

Henderson's Mixed Model Equations

Henderson's mixed model equations (e.g., 1950, 1963, 1975, 1984) simplify for many situations the calculation of $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$. In general form the MME are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \boldsymbol{\Sigma}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

The equations will also be written as $\mathbf{C}\hat{\mathbf{s}} = \mathbf{r}$. Although $\mathbf{R}$ is of order the number of records, $\mathbf{R}$ is usually assumed to be diagonal for single trait analyses, often $\mathbf{I}\sigma_e^2$, and block diagonal (blocks of order of number of traits) for multiple trait analyses, so that calculations with $\mathbf{R}^{-1}$ are easy. Henderson et al. (1959) proved the $\hat{\boldsymbol{\beta}}$ from these equations are BLUE as from generalized least-squares and Henderson (1963) proved the $\hat{\mathbf{u}}$ are BLUP.

Bayesian Variance Components Model

Prior Distributions

To fully specify the model for estimates of variance components additional assumptions must be made. First, the prior distributions for the effects in the model must be determined. These programs were developed using a "flat" prior distribution for the "fixed" effects, that is, there is no prior knowledge about these effects. Next, the random effects are assumed to be normally distributed. For the genetic effects there will be an additional assumption of a known covariance structure among those random effects corresponding to the relationship matrix. Finally, the residual effects are assumed to distributed normally. These assumptions are the same as those used with most likelihood based methods. In addition, these assumptions result in BLUE

and BLUP solutions for fixed and random effects when variance components are known (Gianola and Fernando,1986; Gianola, Im and Macedo, 1990).

The animal and second animal effects are assumed to have non-zero covariances that can be estimated for all trait combinations. Covariances among uncorrelated random effects specified for different traits may be non-zero only if those effects are coded in the same column of the original data set. As seen in Chapter 2, the user can specify that some of these covariances are to be restricted to zero.

The family of prior distributions for the (co)variance components are chosen mainly for computational simplicity. The inverted Wishart (**IW**) distribution is used for (co)variances. The Wishart density describes the distribution of sums of squares and cross-products of standard normal random variables (**RV**s) (Odell and Fieveson, 1966), and if **X** is a Wishart RV, then $\mathbf{X}^{-1}$ is an IW RV. In the univariate case, this corresponds to an inverted Chi-square distribution. Although other distibutions could be used, the Gibbs sampling algorithm would be more complex.

If **T** is distibuted as an IW variable, i.e., $\mathbf{T} \sim IW(\mathbf{V},\nu)$, then the form of the distribution is

$$f(\mathbf{T}|\nu,\mathbf{V}) = K(\nu,\mathbf{V}^{-1})|\mathbf{T}|^{-\frac{1}{2}(\nu+m+1)} \exp\left(\mathrm{tr}\left(-\tfrac{1}{2}\mathbf{V}^{-1}\mathbf{T}^{-1}\right)\right),$$

$$\mathbf{T},\mathbf{V} > 0; \nu > m+1,$$

where,

$$K\left(\nu,\mathbf{V}^{-1}\right) = \left|\mathbf{V}^{-1}\right|^{(1/2)\nu}\left(2^{(1/2)\nu m}\,\pi^{(1/4)m(m-1)}\prod_{j=1}^{m}\Gamma\left(\tfrac{\nu-j+1}{2}\right)\right)^{-1}$$

(Johnson and Kotz, 1972). The parameter ν is an integer variable and is referred to as the shape parameter, corresponding to the degrees of freedom of the corresponding Wishart RV and representing the degree of certainty for the prior distribution, and the matrix **V** describes the variance-covariance structure of the variables. Finally, m is the number of correlated random vectors (dimension of **V**). The mean of **T** is $\mathbf{V}^{-1}/(\nu - m - 1)$ (Johnson and Kotz, 1972). In the MTGSAM programs **V** is calculated such that the expected value of prior density of the genetic (co)variance matrix is equal to the value entered interactively. For example, for the genetic (co)variance matrix, the mean value specified by the user is $\mathbf{G}_0$, then $\mathbf{V}^{-1} = \nu^*\mathbf{G}_0$, where $\nu^* = \nu - m - 1$.

Some previous Gibbs sampling work has been done using flat prior distributions for the variance components (Wang et al., 1994, Jensen et al., 1995). Work by Hobert (1994) indicates that the joint distribution does not always exist when flat priors are used for the variance components, resulting in spurious results from the Gibbs sampler. Further work needs to be done to determine when flat priors can safely be used for variance components.

Joint Posterior Density

The joint posterior desity can be written as the product of the prior and conditional densities previously described. The joint density of the parameters given the data and the prior information is:

$$
\begin{aligned}
&f\left(\boldsymbol{\beta},\mathbf{u},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{R}|\mathbf{y},\nu_g,\mathbf{G}_0,\nu_{d_1},\mathbf{D}_{0_1},\ldots,\nu_{d_\gamma}\mathbf{D}_{0_\gamma},\nu_{r_1},\mathbf{R}_{0_1}\ldots,\nu_{r_\rho},\mathbf{R}_{0_\rho}\right)\\
&\quad\propto f\left(\boldsymbol{\beta},\mathbf{u},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{R},\mathbf{y}|\nu_g,\mathbf{G}_0,\nu_{d_1},\mathbf{D}_{0_1},\ldots,\nu_{d_\gamma}\mathbf{D}_{0_\gamma},\nu_{r_1},\mathbf{R}_{0_1}\ldots,\nu_{r_\rho},\mathbf{R}_{0_\rho}\right)\\
&\quad\propto f\left(\mathbf{y}|\boldsymbol{\beta},\mathbf{u},\mathbf{R}\right)\times f\left(\boldsymbol{\beta}\right)\times f\left(\mathbf{u}|\boldsymbol{\Sigma}\right)\times f\left(\mathbf{G}|\nu_g,\mathbf{G}_0\right)\times\prod_{i=1}^{\gamma}\left[f\left(\mathbf{D}_i|\nu_{d_i},\mathbf{D}_{0_i}\right)\right]\times\prod_{i=1}^{\rho}\left[f\left(\mathbf{R}_i|\nu_{r_i},\mathbf{R}_{0_i}\right)\right]\\
&\quad\propto\left|\mathbf{R}\right|^{-1/2}\times\exp\left\{-\frac{1}{2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}\right)'\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}\right)\right\}\\
&\qquad\times\left|\boldsymbol{\Sigma}\right|^{-1/2}\times\exp\left\{-\frac{1}{2}\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}\right\}\times\left|\mathbf{G}\right|^{-\frac{1}{2}\left(\nu_g+t+1\right)}\times\exp\left(\operatorname{tr}\left(-\frac{\nu_g^*}{2}\mathbf{G}_0\mathbf{G}^{-1}\right)\right)\\
&\qquad\times\prod_{i=1}^{\gamma}\left[\left|\mathbf{D}_i\right|^{-\frac{1}{2}\left(\nu_{d_i}+m_{d_i}+1\right)}\times\exp\left(\operatorname{tr}\left(-\frac{\nu_{d_i}^*}{2}\mathbf{D}_{0_i}\mathbf{D}_i^{-1}\right)\right)\right]\\
&\qquad\times\prod_{i=1}^{\rho}\left[\left|\mathbf{R}_i\right|^{-\frac{1}{2}\left(\nu_{r_i}+m_{r_i}+1\right)}\times\exp\left(\operatorname{tr}\left(-\frac{\nu_{r_i}^*}{2}\mathbf{R}_{0_i}\mathbf{R}_i^{-1}\right)\right)\right].
\end{aligned}
$$

[3.1]

Next, several direct product and direct sum results are presented that are used to simplify the form of the joint density (from Searle, 1982):

$$\left|\mathbf{A}\oplus\mathbf{B}\right|=\left|\mathbf{A}\right|\left|\mathbf{B}\right|$$

$$\left|\mathbf{A}_{p\times p}\otimes\mathbf{B}_{m\times m}\right|=\left|\mathbf{A}\right|^m\left|\mathbf{B}\right|^p$$

$$\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}=\mathbf{u}'\left(\bigoplus_{i=0}^{\gamma}\boldsymbol{\Sigma}_i^{-1}\right)\mathbf{u}=\sum_{i=0}^{\gamma}\mathbf{u}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{u}_i$$

Applying these results,

$$
\begin{aligned}
&f\left(\boldsymbol{\beta},\mathbf{u},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{R}|\mathbf{y},\nu_g,\mathbf{G}_0,\nu_{d_1},\mathbf{D}_{0_1},\ldots,\nu_{d_\gamma}\mathbf{D}_{0_\gamma},\nu_{r_1},\mathbf{R}_{0_1}\ldots,\nu_{r_\rho},\mathbf{R}_{0_\rho}\right)\\
&\propto|\mathbf{R}|^{-1/2}\times\exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})\right\}\\
&\quad\times\prod_{i=0}^{\gamma}|\boldsymbol{\Sigma}_i|^{-1/2}\times\exp\left\{-\frac{1}{2}\sum_{i=0}^{\gamma}\mathbf{u}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{u}_i\right\}\times|\mathbf{G}|^{-\frac{1}{2}(\nu_g+t+1)}\times\exp\left(\operatorname{tr}\left(-\frac{\nu_g^*}{2}\mathbf{G}_0\mathbf{G}^{-1}\right)\right)\\
&\quad\times\prod_{i=1}^{\gamma}\left[|\mathbf{D}_i|^{-\frac{1}{2}(\nu_{d_i}+m_{d_i}+1)}\times\exp\left(\operatorname{tr}\left(-\frac{\nu_{d_i}^*}{2}\mathbf{D}_{0_i}\mathbf{D}_i^{-1}\right)\right)\right]\\
&\quad\times\prod_{i=1}^{\rho}\left[|\mathbf{R}_i|^{-\frac{1}{2}(\nu_{r_i}+m_{r_i}+1)}\times\exp\left(\operatorname{tr}\left(-\frac{\nu_{r_i}^*}{2}\mathbf{R}_{0_i}\mathbf{R}_i^{-1}\right)\right)\right]\\
&\propto|\mathbf{R}|^{-1/2}\times\exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})\right\}\\
&\quad\times|\mathbf{G}|^{-n/2}\times\exp\left\{-\frac{1}{2}\mathbf{u}_0'(\mathbf{G}^{-1}\otimes\mathbf{A}^{-1})\mathbf{u}_0\right\}\times|\mathbf{G}|^{-\frac{1}{2}(\nu_g+t+1)}\times\exp\left(\operatorname{tr}\left(-\frac{\nu_g^*}{2}\mathbf{G}_0\mathbf{G}^{-1}\right)\right)\\
&\quad\times\prod_{i=1}^{\gamma}|\mathbf{D}_i|^{-r_i/2}\times\exp\left\{-\frac{1}{2}\mathbf{u}_i'(\mathbf{D}_i^{-1}\otimes\mathbf{I}_{r_i})\mathbf{u}_i\right\}\times|\mathbf{D}_i|^{-\frac{1}{2}(\nu_{d_i}+m_{d_i}+1)}\times\exp\left(\operatorname{tr}\left(-\frac{\nu_{d_i}^*}{2}\mathbf{D}_{0_i}\mathbf{D}_i^{-1}\right)\right)\\
&\quad\times\prod_{i=1}^{\rho}\left[|\mathbf{R}_i|^{-\frac{1}{2}(\nu_{r_i}+m_{r_i}+1)}\times\exp\left(\operatorname{tr}\left(-\frac{\nu_{r_i}^*}{2}\mathbf{R}_{0_i}\mathbf{R}_i^{-1}\right)\right)\right].
\end{aligned}
\qquad [3.2]
$$

Considering one subvector, $\mathbf{u}_i$, this can be further partitioned into m_{d_i} subvectors as follows:

$\mathbf{u}_i'=\left[\mathbf{u}_{i_1}'\quad\mathbf{u}_{i_2}'\quad\cdots\quad\mathbf{u}_{i_{m_{d_i}}}'\right]$. Finally, define

$$\mathbf{G}^{-1}=\begin{bmatrix}g^{11}&g^{12}&\cdots&g^{1t}\\g^{21}&g^{22}&\cdots&g^{2t}\\\vdots&\vdots&\ddots&\vdots\\g^{t1}&g^{t2}&\cdots&g^{tt}\end{bmatrix},\text{ and}$$

$$\mathbf{S}_0 = \begin{bmatrix} \mathbf{u}_{0_1}'\mathbf{A}^{-1}\mathbf{u}_{0_1} & \mathbf{u}_{0_1}'\mathbf{A}^{-1}\mathbf{u}_{0_2} & \cdots & \mathbf{u}_{0_1}'\mathbf{A}^{-1}\mathbf{u}_{0_t} \\ \mathbf{u}_{0_2}'\mathbf{A}^{-1}\mathbf{u}_{0_1} & \mathbf{u}_{0_2}'\mathbf{A}^{-1}\mathbf{u}_{0_2} & \cdots & \mathbf{u}_{0_2}'\mathbf{A}^{-1}\mathbf{u}_{0_t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_{0_t}'\mathbf{A}^{-1}\mathbf{u}_{0_1} & \mathbf{u}_{0_t}'\mathbf{A}^{-1}\mathbf{u}_{0_2} & \cdots & \mathbf{u}_{0_t}'\mathbf{A}^{-1}\mathbf{u}_{0_t} \end{bmatrix}.$$

Then $\mathbf{u}_i\boldsymbol{\Sigma}_i^{-1}\mathbf{u}_i$ can be rewritten as follows for i = 0:

$$\begin{aligned} \mathbf{u}_0\boldsymbol{\Sigma}_0^{-1}\mathbf{u}_0 &= \begin{bmatrix} \mathbf{u}_{0_1}' & \mathbf{u}_{0_2}' & \cdots & \mathbf{u}_{0_t}' \end{bmatrix}\left[\mathbf{G}^{-1}\otimes\mathbf{A}^{-1}\right]\begin{bmatrix} \mathbf{u}_{0_1} \\ \mathbf{u}_{0_2} \\ \vdots \\ \mathbf{u}_{0_t} \end{bmatrix} \\ &= g^{11}\mathbf{u}_{0_1}'\mathbf{A}^{-1}\mathbf{u}_{0_1} + g^{12}\mathbf{u}_{0_1}'\mathbf{A}^{-1}\mathbf{u}_{0_2} + \cdots + g^{21}\mathbf{u}_{0_2}'\mathbf{A}^{-1}\mathbf{u}_{0_1} \\ &\quad + g^{22}\mathbf{u}_{0_2}'\mathbf{A}^{-1}\mathbf{u}_{0_2} + \cdots + g^{tt}\mathbf{u}_{0_t}'\mathbf{A}^{-1}\mathbf{u}_{0_t} \\ &= \sum_{i=1}^{t}\sum_{j=1}^{t} g^{ii}\mathbf{u}_{0_i}'\mathbf{A}^{-1}\mathbf{u}_{0_i} \\ &= \mathrm{tr}\left(\mathbf{S}_0\mathbf{G}^{-1}\right). \end{aligned}$$

Next, define

$$\mathbf{D}_i^{-1} = \begin{bmatrix} d_i^{11} & d_i^{12} & \cdots & d_i^{1,m_{d_i}} \\ d_i^{22} & d_i^{21} & \cdots & d_i^{2,m_{d_i}} \\ \vdots & \vdots & \ddots & \vdots \\ d_i^{m_{d_i},1} & d_i^{m_{d_i},2} & \cdots & d_i^{m_{d_i},m_{d_i}} \end{bmatrix}, \text{ and}$$

$$\mathbf{S}_i = \begin{bmatrix} \mathbf{u}_{i_1}'\mathbf{u}_{i_1} & \mathbf{u}_{i_1}'\mathbf{u}_{i_2} & \cdots & \mathbf{u}_{i_1}'\mathbf{u}_{i_{m_{d_i}}} \\ \mathbf{u}_{i_2}'\mathbf{u}_{i_1} & \mathbf{u}_{i_2}'\mathbf{u}_{i_2} & \cdots & \mathbf{u}_{i_2}'\mathbf{u}_{i_{m_{d_i}}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{u}_{i_{m_{d_i}}}'\mathbf{u}_{i_1} & \mathbf{u}_{i_{m_{d_i}}}'\mathbf{u}_{i_2} & \cdots & \mathbf{u}_{i_{m_{d_i}}}'\mathbf{u}_{i_{m_{d_i}}} \end{bmatrix}.$$

Then, $\mathbf{u}_i\boldsymbol{\Sigma}_i^{-1}\mathbf{u}_i$ can be rewritten as follows for i > 0:

$$\begin{aligned}
\mathbf{u}_i\boldsymbol{\Sigma}_i^{-1}\mathbf{u}_i &= \mathbf{u}_i\left(\mathbf{D}_i \otimes \mathbf{I}_{n_i}\right)\mathbf{u}_i \\
&= d_i^{11}\mathbf{u}'_{i_1}\mathbf{u}_{i_1} + d_i^{12}\mathbf{u}'_{i_1}\mathbf{u}_{i_2} + \cdots + d_i^{21}\mathbf{u}'_{i_2}\mathbf{u}_{i_1} \\
&\quad + d_i^{22}\mathbf{u}'_{i_2}\mathbf{u}_{i_2} + \cdots + d_i^{m_{d_i},m_{d_i}}\mathbf{u}'_{i_{m_{d_i}}}\mathbf{u}_{i_{m_{d_i}}} \\
&= \sum_{j=1}^{m_{d_i}}\sum_{k=1}^{m_{d_i}} d_i^{jk}\mathbf{u}'_{i_j}\mathbf{u}_{i_k} \\
&= \mathrm{tr}\left(\mathbf{S}_i\mathbf{D}_i^{-1}\right).
\end{aligned}$$

Similarly, let $\mathbf{e} = \mathbf{P}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})$, such that $\mathbf{e}' = \left[\mathbf{e}'_1\,\mathbf{e}'_2\cdots\mathbf{e}'_\rho\right]$, where $\mathbf{e}_i$ is the vector of residuals for block i of the residual (co)variance matrices and **P** is a permutation matrix that simply reorders the elements of the vector so that the residuals are ordered by trait within animal within residual block. The matrix **P** is simply an identity matrix with the columns (or rows) reordered; one property of a permutation matrix is **PP'** = **P'P** = **I**. Then,

$$\mathrm{var}(\mathbf{e}) = \tilde{\mathbf{R}} = \mathbf{P}\mathbf{R}\mathbf{P}',$$

$$\mathbf{P}'\tilde{\mathbf{R}}\mathbf{P} = \mathbf{P}'\mathbf{P}\mathbf{R}\mathbf{P}'\mathbf{P} = \mathbf{I}\mathbf{R}\mathbf{I} = \mathbf{R},$$

$$|\mathbf{R}| = \left|\mathbf{P}'\tilde{\mathbf{R}}\mathbf{P}\right| = |\mathbf{P}'|\left|\tilde{\mathbf{R}}\right||\mathbf{P}| = \left|\tilde{\mathbf{R}}\right||\mathbf{P}'||\mathbf{P}| = \left|\tilde{\mathbf{R}}\right||\mathbf{P}'\mathbf{P}| = \left|\tilde{\mathbf{R}}\right|, \text{and}$$

$$|\mathbf{R}| = \left|\tilde{\mathbf{R}}\right| = \left|\left(\bigoplus_{i=1}^{\rho}\left[\mathbf{R}_i^{-1}\otimes\mathbf{I}_{n_i}\right]\right)\right| = \prod_{i=1}^{\rho}\left|\mathbf{R}_i^{-1}\otimes\mathbf{I}_{n_i}\right| = \prod_{i=1}^{\rho}\left|\mathbf{R}_i^{-1}\right|^{n_i}.$$

The number of traits represented in $\mathbf{e}_i$ is t_i and the number of animals represented is n_i. It is assumed that a residual effect is present for all traits contained in a block if an animal is represented in that block, although an animal is not assumed to be represented in all residual blocks. The impact of missing traits on this requirement will be discussed in the next section on full conditional distributions. Define $\mathbf{e}_{i,j}$ as the residuals for the traits in block i and animal j and $e_{i,j,k}$ as element k of $\mathbf{e}_{i,j}$. Further define

$$\mathbf{R}_i^{-1} = \begin{bmatrix} r_i^{11} & r_i^{12} & \cdots & r_i^{1,t_i} \\ r_i^{21} & r_i^{22} & \cdots & r_i^{2,t_i} \\ \vdots & \vdots & \ddots & \vdots \\ r_i^{t_i,1} & r_i^{t_i,2} & \cdots & r_i^{t_i,t_i} \end{bmatrix}, \text{ and}$$

$$\mathbf{Q}_i = \begin{bmatrix} \mathbf{e}'_{i,1}\mathbf{e}_{i,1} & \mathbf{e}'_{i,1}\mathbf{e}_{i,2} & \cdots & \mathbf{e}'_{i,1}\mathbf{e}_{i,t_i} \\ \mathbf{e}'_{i,2}\mathbf{e}_{i,1} & \mathbf{e}'_{i,2}\mathbf{e}_{i,2} & \cdots & \mathbf{e}'_{i,2}\mathbf{e}_{i,t_i} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{e}'_{i,t_i}\mathbf{e}_{i,1} & \mathbf{e}'_{i,t_i}\mathbf{e}_{i,2} & \cdots & \mathbf{e}'_{i,t_i}\mathbf{e}_{i,t_i} \end{bmatrix}.$$

Then, $(\mathbf{y}\text{-}\mathbf{X}\boldsymbol{\beta}\text{-}\mathbf{Z}\mathbf{u})'\,\mathbf{R}^{-1}\,(\mathbf{y}\text{-}\mathbf{X}\boldsymbol{\beta}\text{-}\mathbf{Z}\mathbf{u})$ can be rewritten as

$$\begin{aligned}
(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'\,\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}) &= \mathbf{e}'\tilde{\mathbf{R}}^{-1}\mathbf{e} \\
&= \mathbf{e}'\left(\bigoplus_{i=1}^{\rho}\left[\mathbf{R}_i^{-1}\otimes\mathbf{I}_{n_i}\right]\right)\mathbf{e} \\
&= \sum_{i=1}^{\rho}\mathbf{e}'_i\left[\mathbf{R}_i^{-1}\otimes\mathbf{I}_{n_i}\right]\mathbf{e}_i \\
&= \sum_{i=1}^{\rho}\sum_{j=1}^{n_i}\mathbf{e}'_{i,j}\mathbf{R}_i^{-1}\mathbf{e}_{i,j} \\
&= \sum_{i=1}^{\rho}\sum_{j=1}^{n_i}\sum_{k=1}^{t_i}\sum_{l=1}^{t_i} r_i^{k,l}\mathbf{e}'_{i,j,k}\mathbf{e}_{i,j,l} \\
&= \sum_{i=1}^{\rho}\mathrm{tr}\left(\mathbf{Q}_i\mathbf{R}_i^{-1}\right).
\end{aligned}$$

Then finally,

$$\begin{aligned}
&f\left(\boldsymbol{\beta},\mathbf{u},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{R}|\mathbf{y},\nu_g,\mathbf{G}_0,\nu_{d_1},\mathbf{D}_{0_1},\ldots,\nu_{d_\gamma}\mathbf{D}_{0_\gamma},\nu_{r_1},\mathbf{R}_{0_1}\ldots,\nu_{r_\rho},\mathbf{R}_{0_\rho}\right) \\
&\propto |\mathbf{G}|^{-\frac{1}{2}\left(n+\nu_g+t+1\right)}\times\exp\left(-\frac{1}{2}\mathrm{tr}\left(\left(\nu_g^*\mathbf{G}_0+\mathbf{S}_0\right)\mathbf{G}^{-1}\right)\right) \\
&\times\prod_{i=1}^{\gamma}\left[|\mathbf{D}_i|^{-\frac{1}{2}\left(r_i+\nu_{d_i}+m_{d_i}+1\right)}\times\exp\left(-\frac{1}{2}\mathrm{tr}\left(\left(\nu_{d_i}^*\mathbf{D}_{0_i}+\mathbf{S}_i\right)\mathbf{D}_i^{-1}\right)\right)\right] \\
&\times\prod_{i=1}^{\rho}\left[|\mathbf{R}_i|^{-\frac{1}{2}\left(n_i+\nu_{r_i}+m_{r_i}+1\right)}\times\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\left(\nu_{r_i}^*\mathbf{R}_{0_i}+\mathbf{Q}_i\right)\mathbf{R}_i^{-1}\right)\right\}\right].
\end{aligned}$$

[3.3]

Full Conditional Densities

The full conditional densities required for GS can be derived from the different versions of the joint posterior density, i.e., [3.1], [3.2], and [3.3], by treating the known parameters as

constants and reorganizing the remaining variables into the form of the kernel of a recognized density.

Fixed and random effects

First, the conditional distribution of the "solutions" of the mixed model equations (i.e., the "fixed" and random effects) will be obtained. A useful result from the form of the mixed model equations is:

$$\begin{aligned}
\mathbf{s}'\mathbf{C}\mathbf{s} &= \begin{bmatrix}\boldsymbol{\beta}' & \mathbf{u}'\end{bmatrix}\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\boldsymbol{\Sigma}^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta} \\ \mathbf{u}\end{bmatrix} \\
&= \begin{bmatrix}\boldsymbol{\beta}' & \mathbf{u}'\end{bmatrix}\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta} \\ \mathbf{u}\end{bmatrix} + \begin{bmatrix}\boldsymbol{\beta}' & \mathbf{u}'\end{bmatrix}\begin{bmatrix}\mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta} \\ \mathbf{u}\end{bmatrix} \\
&= \mathbf{s}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}\mathbf{s} + \mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}, \text{ where} \\
\mathbf{W} &= \begin{bmatrix}\mathbf{X} & \mathbf{Z}\end{bmatrix}.
\end{aligned}$$

Then, the conditional distribution of the solutions, $\mathbf{s} = [\boldsymbol{\beta}' \; \mathbf{u}']'$, can be written as

$$\begin{aligned}
&f\left(\mathbf{s}|\mathbf{R},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{y}\right) \\
&\propto \exp\left\{-\frac{1}{2}\left(\mathbf{y}-\mathbf{W}\mathbf{s}\right)'\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{W}\mathbf{s}\right)\right\} \times \exp\left\{-\frac{1}{2}\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(-\mathbf{y}'\mathbf{R}^{-1}\mathbf{W}\mathbf{s}-\mathbf{s}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{y}+\mathbf{s}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}\mathbf{s}+\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\mathbf{s}'\mathbf{C}\mathbf{s}-2\mathbf{s}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{y}\right)\right\}
\end{aligned}$$

Constants with respect to **s** are added to complete the square for the quadratic form and the density rewritten as

$$\begin{aligned}
&f\left(\mathbf{s}|\mathbf{R},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{y}\right) \\
&\propto \exp\left\{-\frac{1}{2}\left(\mathbf{s}'\mathbf{C}\mathbf{s}-2\mathbf{s}'\mathbf{W}'\mathbf{R}^{-1}\mathbf{y}\right)\right\} \times \exp\left\{-\frac{1}{2}\left(\mathbf{y}'\mathbf{R}^{-1}\mathbf{W}\mathbf{C}^{-1}\mathbf{W}'\mathbf{R}^{-1}\mathbf{y}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\left(\mathbf{s}-\tilde{\mathbf{s}}\right)'\mathbf{C}\left(\mathbf{s}-\tilde{\mathbf{s}}\right)\right)\right\}, \text{ where} \\
\tilde{\mathbf{s}} &= \mathbf{C}^{-1}\mathbf{W}'\mathbf{R}^{-1}\mathbf{y}.
\end{aligned}$$

This is the kernel of a normal density, and therefore,

$$\mathbf{s}|\mathbf{R},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{y} \sim \mathrm{N}\left(\tilde{\mathbf{s}},\mathbf{C}^{-1}\right).$$

[3.4]

A result for conditional normal distributions

A result that will be useful in obtaining conditional distributions for individual fixed or random effects or for subvectors will be derived. The conditional distribution for a group of random variables can be derived using partitioned matrix results (Searle, 1982) and the form of the conditional normal density (see Searle, 1971 for example). The result will be derived for the special case of the multivariate normal distribution having the form $\mathbf{x} \sim \mathrm{N}(\mathbf{C}^{-1}\mathbf{r},\mathbf{C}^{-1})$. The derivation will be done for the first group of elements in **x**, but this is done without loss of generality, because the order of the elements in the vector is arbitarary and can be changed using a permutation matrix to reorder the elements in the mean and variance. Consider the partition of **C** and **r** such that $\mathbf{r}' = \begin{bmatrix}\mathbf{r}_1' & \mathbf{r}_2'\end{bmatrix}$ and **C** is partitioned such that the leading subdiagonal is of the same order as $\mathbf{r}_1'$. Then,

$$\mathbf{C} = \begin{bmatrix}\mathbf{Q} & \mathbf{S} \\ \mathbf{S}' & \mathbf{T}\end{bmatrix}, \text{and } \mathbf{C}^{-1} = \begin{bmatrix}\mathbf{E} & \mathbf{F} \\ \mathbf{F}' & \mathbf{H}\end{bmatrix}.$$

From Searle (1982):

$$\begin{aligned}
\mathbf{H} &= \left(\mathbf{T} - \mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}\right)^{-1}, \\
\mathbf{E} &= \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{S}\left(\mathbf{T} - \mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}\right)^{-1}\mathbf{S}'\mathbf{Q}^{-1} \\
&= \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1}, \text{and} \\
\mathbf{F} &= -\mathbf{Q}^{-1}\mathbf{S}\left(\mathbf{T} - \mathbf{S}'\mathbf{Q}^{-1}\mathbf{S}\right)^{-1} \\
&= -\mathbf{Q}^{-1}\mathbf{S}\mathbf{H}.
\end{aligned}$$

Then, the mean of the distribution is

$$
\begin{aligned}
\mathrm{E}(\mathbf{x}) &= \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{F}' & \mathbf{H} \end{bmatrix} \mathbf{r} \\
&= \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{F}' & \mathbf{H} \end{bmatrix} \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{E}\mathbf{r}_1 + \mathbf{F}\mathbf{r}_2 \\ \mathbf{F}'\mathbf{r}_1 + \mathbf{H}\mathbf{r}_2 \end{bmatrix} \\
&= \begin{bmatrix} \left(\mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1}\right)\mathbf{r}_1 - \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{r}_2 \\ -\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1}\mathbf{r}_1 + \mathbf{H}\mathbf{r}_2 \end{bmatrix}.
\end{aligned}
$$

Next, the form of the conditional normal is required. Based on Searle (1971), if

$$
\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathrm{N}\left[\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \right], \text{then,}
$$

$$
\mathbf{x}_1|\mathbf{x}_2 \sim \mathrm{N}\left[\boldsymbol{\mu}_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2), (\mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21})\right]. \qquad [3.5]
$$

Finally,

$$
\mathbf{x}_1|\mathbf{x}_2 \sim \mathrm{N}(\mu_1^*, \mathbf{V}_{11}^*),
$$

where

$$
\begin{aligned}
\mu_1^* &= \boldsymbol{\mu}_1 + \mathbf{F}\mathbf{H}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\
&= \left(\left(\mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1}\right)\mathbf{r}_1 - \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{r}_2\right) \\
&\quad - \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{H}^{-1}\left(\mathbf{x}_2 - \left(-\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1}\mathbf{r}_1 + \mathbf{H}\mathbf{r}_2\right)\right) \\
&= \mathbf{Q}^{-1}\mathbf{r}_1 + \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1}\mathbf{r}_1 - \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{r}_2 \\
&\quad - \mathbf{Q}^{-1}\mathbf{S}\mathbf{x}_2 - \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1}\mathbf{r}_1 + \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{r}_2 \\
&= \mathbf{Q}^{-1}\mathbf{r}_1 - \mathbf{Q}^{-1}\mathbf{S}\mathbf{x}_2
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{V}_{11}^* &= \mathbf{E} - \mathbf{F}\mathbf{H}^{-1}\mathbf{F}' \\
&= \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1} - \left(-\mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\right)\mathbf{H}^{-1}\left(-\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1}\right) \\
&= \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{S}\mathbf{H}\mathbf{S}'\mathbf{Q}^{-1} \\
&= \mathbf{Q}^{-1}.
\end{aligned}
$$

If this result is applied to the vector of fixed and random effects,

$$\mathbf{s}_1|\mathbf{s}_2 \sim \mathrm{N}\left(\mathbf{C}_{11}^{-1}\left(\mathbf{r}_1 - \mathbf{C}_{12}\mathrm{s}_2\right), \mathbf{C}_{11}^{-1}\right), \qquad [3.6]$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}.$$

The conditional distributions for the fixed and random effects will also be considered individually.

Fixed effects

First, considering only terms that involve $\boldsymbol{\beta}$, the full conditional density of the fixed effects is:

$$\begin{aligned}
&\mathrm{f}\left(\boldsymbol{\beta}|\mathbf{u}, \mathbf{R}, \mathbf{y}\right) \\
&\propto \exp\left\{-\frac{1}{2}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}\right)' \mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(-\mathbf{y}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\mathbf{R}^{-1}\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} + \mathbf{u}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta}\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}'\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{Z}\mathbf{u}\right)\right)\right\}
\end{aligned}$$

Constants (with respect to $\boldsymbol{\beta}$) are added to complete the quadratic from, and the density can be written as:

$$
\begin{aligned}
&f(\boldsymbol{\beta}|\mathbf{u},\mathbf{R},\mathbf{y}) \\
&\propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}'\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Z}\mathbf{u})\right)\right\} \\
&\quad\times\exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Z}\mathbf{u})\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\left(\boldsymbol{\beta}-\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Z}\mathbf{u})\right)'\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)\left(\boldsymbol{\beta}-\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Z}\mathbf{u})\right)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}\right)'\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}\right)\right)\right\}, \\
&\text{where}\ \tilde{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Z}\mathbf{u}).
\end{aligned}
$$

This is the kernel of a normal density, and therefore,

$$\boldsymbol{\beta}|\mathbf{u},\mathbf{R},\mathbf{y} \sim \mathrm{N}\left(\tilde{\boldsymbol{\beta}},\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\right).$$

Note that this result can also be obtained by applying the special form of the conditional normal, [3.6], to the form of the full conditional distribution of the fixed and random effects, [3.4].

Random effects

Next, considering only terms that involve **u**, the full conditional density of the random effects is:

$$
\begin{aligned}
&f(\mathbf{u}|\boldsymbol{\beta},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{R},\mathbf{y}) \\
&\quad\propto \exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})\right\}\times\exp\left\{-\frac{1}{2}\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}\right\} \\
&\quad\propto \exp\left\{-\frac{1}{2}\left(\mathbf{u}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{u} + 2\mathbf{u}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\boldsymbol{\beta} - 2\mathbf{u}'\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}\right)\right\} \\
&\quad\propto \exp\left\{-\frac{1}{2}\left(\mathbf{u}'\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\boldsymbol{\Sigma}^{-1}\right)\mathbf{u} - 2\mathbf{u}'\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right)\right\}.
\end{aligned}
$$

Constants (with respect to **u**) are added to complete the quadratic from, and the density can be written as:

$$
\begin{aligned}
&f\left(\mathbf{u}|\boldsymbol{\beta},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{R},\mathbf{y}\right) \\
&\propto \exp\left\{-\frac{1}{2}\left(\mathbf{u}'\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\boldsymbol{\Sigma}^{-1}\right)\mathbf{u}-2\mathbf{u}'\mathbf{Z}'\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right)\right\} \\
&\quad\times\exp\left\{-\frac{1}{2}\left(\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\mathbf{R}^{-1}\mathbf{Z}\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\boldsymbol{\Sigma}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\left(\mathbf{u}-\tilde{\mathbf{u}}\right)'\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\boldsymbol{\Sigma}^{-1}\right)\left(\mathbf{u}-\tilde{\mathbf{u}}\right)\right)\right\}, \\
&\text{where } \tilde{\mathbf{u}}=\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\boldsymbol{\Sigma}^{-1}\right)^{-1}\mathbf{Z}'\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right).
\end{aligned}
$$

This is also the kernel of a normal density, and therefore,

$$
\mathbf{u}|\boldsymbol{\beta},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{R},\mathbf{y}\sim\mathrm{N}\left(\tilde{\mathbf{u}},\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\boldsymbol{\Sigma}^{-1}\right)^{-1}\right).
$$

[3.7]

Blocking random effects

A blocked Gibbs sampling algorithm is used to generate all genetic effects for an animal as well as associated uncorrelated random effects simultaneously. For example, for dairy cattle data with multiple lactations recorded, often a permanent environmental effect is included to account for non-genetic, animal specific, random effects. The permanent environmental levels would correspond to the animal levels, that is, the level for animal identification and permanent environmental (PE) effect would be coded in the same column of the original data set. The MTGSAM programs would in that case generate the PE effects simultaneously with the genetic effects. In general, the programs block any uncorrelated random effect that is coded in the same column as an animal or second animal effect with the genetic effects when generating new values, so that al effects in the block are generated simultaneously. Generating correlated variables in blocks will often increase the mixing rate of the Gibbs sampler and reduce the correlations among samples drawn. This method seems to help reduce correlations among the samples drawn, but does not completely eliminate the highly correlated samples obtained using GS (Van Tassell, Casella, and Pollak, 1994; Liu, Wong and Kong, 1994). A correlation can exist between variables that are assumed to be statistically uncorrelated caused by the data structure (i.e., through the least

squares part of the equations). Intuitively, if the animal and PE effects are generated individually then the range of one effect, say PE, is limited by the current value of the genetic effect. This in turn leads to the animal effect being limited by the current value of the PE, resulting in increased correlation among sequential rounds of values in the Gibbs sampler, which reduces the efficiency and convergence of the Gibbs sampler. In models with maternal genetic and permanent environmental effects, those effects are blocked together; that is the maternal permanent environmental effect is blocked with the maternal genetic effect because those effects are correlated due to the data structure. Previous results for a maternal effects model applied to Simmental weaning weight data support this concept, although the difference between blocked and scalar algorithms were relatively small (Van Tassell, Casella, and Pollak, 1994).

In order to derive the form of full conditional distribution for the block of effects additional (still more!) definitions are needed. Let $\mathbf{P}$ be a permutation matrix such that

$$\mathbf{Pu} = \begin{bmatrix} \mathbf{P}_i\mathbf{u} \\ \mathbf{P}_{-i}\mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_i \\ \mathbf{u}_{-i} \end{bmatrix},$$

where $\mathbf{u}_i$ are the random effects in the block, and $\mathbf{u}_{-i}$ are the remaining random effects. Then applying the special form of the conditional normal distribution that was derived, i.e., [3.6], to the full conditional distribution of the random effects in [3.7] results in the full conditional distribution of the block of random effects:

$$\mathbf{u}_i|\mathbf{u}_{-i},\boldsymbol{\beta},\mathbf{G},\mathbf{D}_1,\mathbf{D}_2,\ldots,\mathbf{D}_k,\mathbf{R},\mathbf{y} \sim N\left(\tilde{\mathbf{u}}_i,\left(\mathbf{P}_i\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\boldsymbol{\Sigma}^{-1}\right)\mathbf{P}_i'\right)^{-1}\right), \qquad [3.8]$$

where

$$\tilde{\mathbf{u}}_i = \left(\mathbf{P}_i\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\boldsymbol{\Sigma}^{-1}\right)\mathbf{P}_i'\right)^{-1}\left(\mathbf{P}_i\mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})-\mathbf{P}_i\left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\boldsymbol{\Sigma}^{-1}\right)\mathbf{P}_{-i}'\mathbf{u}_{-i}\right).$$

Although this form appears quite complex, it is actually fairly simple. The variance matrix is comprised of the appropriate elements of the coefficient matrix, the mean is a function of that matrix and the right hand sides for the same effects adjusted for the off-diagonal elements of the effects not generated in that block. The adjustments to the right hand sides are based on the rows of the coefficient matrix for those blocked elements with the columns removed for the elements

included in that block. The remaining matrix is multiplied by the vector of effects not included in the block and subtracted from the appropriate element.

Residual effects with missing data

To allow for missing traits, the residual effects for missing traits must be generated in order to calculate quadratic forms for residual effects and for generating residual (co)variance matrices. The missing residuals are generated using the form of the conditional normal distribution, i.e., the residuals are calculated for the observed traits and the missing residuals are generated using the Gibbs sampler based on the current values of the (co)variances for that block. In order to specify the form of the full conditional distribution of the missing residual effects, assume, without loss of generality, that the missing traits occur in the first variables of a block. That is, the vector of residuals can be partitioned as

$$\mathbf{e}_{i,j} = \begin{bmatrix} \mathbf{e}_m \\ \mathbf{e}_o \end{bmatrix},$$

where $\mathbf{e}_m$ is the sub-vector of missing residuals and $\mathbf{e}_o$ is the sub-vector of residuals for observed traits. For block i of residual effects, the residuals are assumed to be distributed normally, specifically,

$$\mathbf{e}_i | \mathbf{R}_i \sim N(\mathbf{0}, \mathbf{R}_i).$$

Let

$$\mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{mm} & \mathbf{R}_{mo} \\ \mathbf{R}_{om} & \mathbf{R}_{oo} \end{bmatrix},$$

then, applying the result for the general conditional normal distribution,

$$\mathbf{e}_m | \mathbf{e}_o, \mathbf{R}_i \sim N\left(\mathbf{R}_{mo}\mathbf{R}_{oo}^{-1}\mathbf{e}_o, \mathbf{R}_{mm} - \mathbf{R}_{mo}\mathbf{R}_{oo}^{-1}\mathbf{R}_{om}\right). \quad [3.9]$$

Finally, the full conditional distributions of the variance components are derived.

Genetic (co)variance matrix

Using the final form of the joint posterior distribution,

$$f\left(\mathbf{G} | \mathbf{u}, \mathbf{G}_0, \nu_g\right) \propto |\mathbf{G}|^{-\frac{1}{2}\left(n+\nu_g+t+1\right)} \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\left(\nu_g^* \mathbf{G}_0 + \mathbf{S}_0\right)\mathbf{G}^{-1}\right)\right).$$

This is the kernel of an IW density, specifically,

$$\mathbf{G}|\mathbf{u},\mathbf{G}_0,\nu_g \sim \text{IW}\left(\left(\nu_g^*\mathbf{G}_0+\mathbf{S}_0\right)^{-1},n+\nu_g\right).$$

[3.10]

Uncorrelated random effect (co)variance matrices

The full conditional distribution for the matrix of (co)variances for each group of uncorrelated random effects is

$$f\left(\mathbf{D}_i|\mathbf{u},\mathbf{D}_{0_i},\nu_{d_i}\right)\propto\left|\mathbf{D}_i\right|^{-\frac{1}{2}\left(r_i+\nu_{d_i}+m_{d_i}+1\right)}\times\exp\left(-\frac{1}{2}\text{tr}\left(\left(\nu_{d_i}^*\mathbf{D}_{0_i}+\mathbf{S}_i\right)\mathbf{D}_i^{-1}\right)\right).$$

This also is the kernel of an IW density; specifically,

$$\mathbf{D}_i|\mathbf{u},\mathbf{D}_{0_i},\nu_{d_i}\sim\text{IW}\left(\left(\nu_{d_i}^*\mathbf{D}_{0_i}+\mathbf{S}_i\right)^{-1},r_i+\nu_{d_i}\right).$$

[3.11]

Residual (co)variance matrices

The full conditional distribution for the matrix of (co)variances for each group of residual effects is

$$f\left(\mathbf{R}_i|\mathbf{e}_i,\nu_{r_i},\mathbf{R}_{0_i}\right)\propto\left|\mathbf{R}_i\right|^{-\frac{1}{2}\left(n_i+\nu_{r_i}+m_{r_i}+1\right)}\times\exp\left\{-\frac{1}{2}\text{tr}\left(\left(\nu_{r_i}^*\mathbf{R}_{0_i}+\mathbf{Q}_i\right)\mathbf{R}_i^{-1}\right)\right\}.$$

This also is the kernel of an IW density; specifically,

$$\mathbf{R}_i|\mathbf{e}_i,\nu_{r_i},\mathbf{R}_{0_i}\sim\text{IW}\left(\left(\nu_{r_i}^*\mathbf{R}_{0_i}+\mathbf{Q}_i\right)^{-1},n_i+\nu_{r_i}\right).$$

[3.12]

Implementation of the Gibbs sampler

Based on the full conditional distributions derived, the GS algorithm used can be outlined as:

1. Calculate starting values for all variables.

    a. Means of (co)variances supplied by the user are used for those components.

b. Gauss-Seidel iteration using the starting variance components is used to generate starting values for fixed and random effects. The user specifies maximum number of rounds of iteration and convergence criterion.

2. Generate fixed effects from [3.6].
3. Generate genetic effects and blocked uncorrelated random effects from [3.8].
4. Generate uncorrelated random effects not in a block from [3.6].
5. Calculate residual effects for traits with observations, and generate missing residuals from [3.9].
6. Calculate quadratic forms for genetic effects, $\mathbf{S}_0$ from $\mathbf{u}'_{0_i}\mathbf{A}^{-1}\mathbf{u}_{0_j}$.
7. Generate **G** from [3.10].
8. Calculate quadratic forms for each block of uncorrelated random effects, $\mathbf{S}_i$ from $\mathbf{u}'_{i_j}\mathbf{u}_{i_k}$.
9. Generate each $\mathbf{D}_i$ from [3.11].
10. Calculate quadratic forms for each block of residual effects, $\mathbf{Q}_i$ from $\mathbf{e}'_i\mathbf{e}_j$.
11. Generate each $\mathbf{R}_i$ from [3.12].
12. Repeat steps 2 through 11 (many times!).

Estimation of Parameter Means and Posterior Distributions

Typically, the mean of a parameter is the point estimate of interest. There are two basic methods that are used by MTGSAM, depending on the variable. The first is based on the average of the expected values of the parameter and the second based on the average of the sampled values. For most variables, the expected value of the parameter is used because it is the Rao-Blackwell estimator; i.e., , it is the minimum variance estimator. The algorithm is quite simple; the program determines the total number of samples that will be included in the average, to determine the denominator of the average. The number is simply the total rounds of GS minus the number of rounds discarded in the burn-in phase of the analysis. The variable for the mean is initialized before the GS algorithm starts. Then, for each round of post burn-in GS the expected value of the parameter divided by the denominator is added to the variable for the mean. The division is done as the values are added to reduce the likelihood of overflow errors as the number of values added might be large in some analyses. This is at the price of computational efficiency

as well, as many divisions are done for each mean rather than just one division done when the sum is completed. The expected value used in the calculation correspond to the expected value of the full conditional distribution for that parameter from which the new value is sampled. There are cases where the expected value of the parameter is not known; specifically, for functions of parameters. These include phenotypic variance, correlations, heritabilities (or other fractions of phenotypic variance), and linear combinations of fixed and random effects (i.e., contrasts, estimable functions, and predictible functions). A simple average of the observed values for each of those values are used in that case. For example, to calculate the mean of the difference of two fixed effects, the observed difference would be calculated in each round and that value averaged for all of the post burn-in rounds.

Similarly, a parametric approach can be used for the estimation of the posterior distibution of some parameters. This is done by calculating the average height of the conditional distribution across the range of values for the parameter (Casella and George, 1992). For example to calculate the posterior distibution of a scalar x, and the full conditional distribution includes a vector **y**, then to determine the estimated posterior distribution of f(x) calculate for values of x_i throughtout the range of the parameter

$$\hat{f}(x_i) = \frac{1}{n}\sum_{j=1}^{n} f\left(x_i|\mathbf{y}_j\right),$$

[3.13]

Note that the vectors $\mathbf{y}_j$ correspond to samples from the Gibbs sampler which are assumed to be independent; this typically means that the samples are taken some number of rounds apart in the Gibbs chain. The distance between using samples to calculate the posterior distributions will vary with the data set and the model used.

Because the conditional distribuion for all parameters cannot be written in closed form alternative methods may need to be used. The marginal distribution of individual elements of an inverted Wishart cannot be written in closed form if there is more than a single variance. Recall that the inverted Wishart simplifies to the inverted chi-square with only one random effect. The parametric approach can be used with inverted chi-square variables since the full conditional distribution can be written in closed form. One simple alternative is to generate a histogram of

the samples drawn for a parameter. The precision of distribution estimates obtained using these methods can be dramatically different; it takes **many** fewer points from the Gibbs sampler to estimate the distribution using the parametric approach. There are more sophisticated approaches than simple histograms available for non-parametric density estimation. One example is the average shifted histogram (**ASH**) algorithm described by Scott (1992). This method uses a flexible weighting algorithm to average the height of neighboring cells in the histogram to smooth the density estimate.

Variance Components

Posterior Mean

The MTGSAM posterior mean estimate for variance components is based on the expected value of the IW RV. Recall that if $\mathbf{T} \sim IW(\mathbf{V},\nu)$, then $E(\mathbf{T}) = \mathbf{V}^{-1}/(\nu - m - 1)$. Therefore, from [3.10], [3.11], and [3.12] the expected values for variance matrices in a given round of Gibbs sampling are calculated as

$$E\left(\mathbf{G}|\mathbf{S}_0,\mathbf{G}_0,\nu_g\right) = \frac{\nu_g^*\mathbf{G}_0 + \mathbf{S}_0}{\left(n + \nu_g - t - 1\right)},$$

$$E\left(\mathbf{D}_i|\mathbf{S}_i,\mathbf{D}_{0_i},\nu_{d_i}\right) = \frac{\nu_{d_i}^*\mathbf{D}_{0_i} + \mathbf{S}_i}{\left(r_i + \nu_{d_i} - m_{d_i} - 1\right)},$$

and

$$E\left(\mathbf{R}_i|\mathbf{Q}_i,\nu_{r_i},\mathbf{R}_{0_i}\right) = \frac{\nu_{r_i}^*\mathbf{R}_{0_i} + \mathbf{Q}_i}{\left(n_i + \nu_{r_i} - m_{r_i} - 1\right)}.$$

The mean of the (co)variance components were calculated as the average of these expected values over the length of the post burn-in chain.

Posterior Distribution

The MTGSAM program does not estimate the posterior distribution for the variance components. At some later date an additional program may be added to assist in obtaining posterior density estimates. Until that time, the ASH programs will be distributed to assist the user in non-parametric density estimation. The data needed to generate the poster distributions for

the (co)variance components is written to units 61, 62, and 63. Information about the format of those files and how to estimate those posterior densities is given in chapter 5.

It is possible to use a parametric density estimate only when there is a singe random effect, i.e., the variance matrix is a scalar. In that case, the full conditional distribution for the variance component is a scaled inverted chi-square distrtibution, which can also be written as an inverted gamma distribution. The form of the inverted gamma (IG(α,γ)) distribution is

$$\mathrm{ig}(x|\alpha,\gamma) = \frac{1}{\Gamma(\alpha)\gamma^{\alpha}}(x)^{-\alpha-1}\exp\left\{\frac{-1}{\gamma x}\right\}.$$

Following the concept of [3.13], the posterior distribution can be estimated as a mixture distribution of the full conditional distributions from which the variance component was drawn. Specifically, for a value, σ_i^2, in the range of the parameter:

$$\begin{aligned}\hat{f}(\sigma_i^2|\mathbf{y}) &= \tfrac{1}{n}\sum_{j=1}^{n}\mathrm{ig}\left(\sigma_i^2|\tfrac{\nu}{2},2s_j\right)\\ &= \tfrac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{\Gamma\left(\frac{\nu}{2}\right)\left(2s_j\right)^{\frac{\nu}{2}}}\left(\sigma_i^2\right)^{-\frac{\nu}{2}-1}\exp\left\{\frac{-1}{2s_j\sigma_i^2}\right\}\right),\end{aligned}$$

where s_j is the scale parameter of the univariate IW variable (i.e., the inverse of the combined value representing the sum of squares and prior information for a variance component), ν is the shape parameter of the IW variable (i.e., the combination of the number of levels of the random effect and the prior distribution shape parameter), and n is the number of samples available to estimate the posterior distribution. The s_j values are assumed to be sampled from far enough apart in the Gibbs sampling chain to be effectively uncorrelated.

Functions of Variance Components

Posterior Mean

The functions of variance components considered by MTGSAM include the phenotypic (co)variances (sum of appropriate genetic, uncorrelated random and residual (co)variances), correlations, and fraction of phenotypic variance accounted for by a particular variance component (e.g., heritability (h^2), fraction due to uncorrelated random effects (c^2), or fraction due to residual

effects (e^2)). Because the conditional distribution of these functions cannot be written in closed form a non-parametric approach is used by MTGSAM to estimate the posterior means for these values. The posterior means for the parameters are estimated as the mean of the functions calculated using the sampled variance components in each post burn-in round of Gibbs sampling.

Posterior Distribution

The MTGSAM program does not estimate the posterior distribution for functions of the variance components. The ASH programs can be used for non-parametric density estimation using the observed values of the functions. The values of those functions of the VCs are not written to the GS files, and, therefore, must be recalculated from the values of the VCs. The values of the VCs are written to unit 61. Information about the format of that file and the ASH algorithm is given in Chapter 5.

Fixed and Random Effects

Posterior Mean

Recall that the full conditional distribution of the fixed and random effects is a normal distribution. The posterior means for those parameters are estimated in MTGSAM as the average of the means of the normal distribution that the parameter is sampled from in each of the post burn-in round of GS. The user decides if the mean of the fixed and random effects is written to files in MTGSAM. If requested, them mean estimates of the covariates are written to unit 71, animal and second animal means to unit 72, and uncorrelated random effect means to unit 73.

Posterior Distribution

The variance of the full conditional distribution of the fixed and random effects is simply the inverse of the corresponding diagonal element of the coefficient matrix. This is true even for the random effects generated in blocks (this can be shown using the form of the conditional normal distribution). Given the mean and the variance, the posterior distribution can be estimated as a mixture distribution of normal distributions. The same technique used for the parametric density estimate for the variance components can be used for the fixed and random effects, which is based on [3.13]. Specifically, for a value, s_i, in the range of the parameter:

$$\hat{f}(s_i|\mathbf{y}) = \tfrac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma_j}\exp\left\{-\frac{(s_i-\mu_j)^2}{2\sigma_j}\right\}\right),$$

where μ_j and σ_j are the mean and standard deviation, respectively, of the full conditional distribution of the fixed or random effect. Again, it is assumed that the samples are taken far enough apart in the GS chain that the parameters are effectively uncorrelated.

If requested, the observed values of the fixed and random effects are written to unit 61 and the mean and variances of the distribution from which the values were sampled are written to unit 62. The information in the file for unit 63 provides information about how to read those data files. Information about the format of those files and how to estimate those posterior densities is given in chapter 5.

Functions of Fixed and Random Effects

The MTGSAM programs consider two different linear cominations of fixed and random effects: those based on a single effect and those based on multiple fixed or random effects. A contrast of a single element may be used if a particular effect is of interest. An example might be a case where the distribution of genetic effects for a select group of animals is wanted, but the amount of information generated for all animals would be prohibitive (i.e., selecting the option for writing sample information for solutions is impracticle). When the contrast contains only one effect the program uses the parametric estimate of the mean of the contrast, i.e., the average of the means of the normal full conditional distributions. In addition, the sample information written to the unit 62 includes the mean and variance of the normal full conditional distributions, which allows for the parametric estimation of the posterior distribution. When the contrast includes multiple effects a non-parametric approach must be used for mean and posterior density estimation. The mean is estimated as the average of the sampled values, which are determined by calculated the linear combination of sampled fixed and random effects specified by the contrast. These values are written to unit 61. In addition, because there is no full conditional distribution for the contrast the mean written to unit 62 is simply the observed value of the contrast and the variance is written as 0.0 so that the user can identify which contrasts can be evaluated using

parametric tools. Chapter 5 describes the format of files for units 61, 62, and 63 as well as discussed estimation of posterior distributions.

Posterior Mean

For single element contrasts the mean is calculated as the average of the normal full conditional distribution for the corresponding effect. For contrasts based on multiple effects, the mean is calculated as the average of the sampled values, which are calculated from the sampled fixed and random effects included in the contrast.

Posterior Distribution

To estimate the posterior distribution of a contrast including a single effect one can use a parametric approach based on the normal full conditional distribution of the fixed or random effect (see the section on posterior distributions for fixed and random effects) or by using a non-parametric approach (such as the ASH algorithm). The posterior distribution for contrasts based on multiple effects must be calculated using some non-parametric method where the observed values of the contrasts are analysed.

CHAPTER FOUR: Computational Strategies for MTGSAM

MTGSAM consists of three programs: MTGSNRM which calculates the non-zero elements of the inverse of the numerator relationship matrix, MTGSPREP which determines the non-zero elements of **W** =[**X Z**] for each animal (these elements are needed to build the MME), and MTGSRUN which implements the Gibbs sampler for a variety of models.

Because only a few changes were made in the MTDFNRM program to create MTGSNRM and because that program follows the method of Quaas (1976), the program techniques employed in that program will not be described here.

MTGSPREP

Most of the MTGSPREP program corresponds to the MTDFPREP program, and much of the following text is taken from Chapter 5 of the MTDFREML manual. Some of the modifications needed for MTGSPREP will also be described.

Basically, MTGSPREP forms the part of the MME that is independent of **G** and **R** used in each round, i.e., **W**=[**X Z**] and **y**. Note that in MTDFREML both the data and the covariates are calculated as deviations from their respective means. As with that program, MTGSPREP can be modified to produce either original or deviated values for each of these variables. To locate the appropriate section of the program to make modifications, search for the phrase *COVARIATE DEVIATION* or *DATA DEVIATION*, for the covariates and data sections of the program, respectively. Deviated values for covariates and data will be used for the example presented later in this chapter.

With appropriate modification of the include file (GSPARAM.FOR), this program can fit any number of fixed effects (both discrete and continuous) and random effects in addition to the required animal effect. Models can be different for each trait and missing observations are permitted.

MTGSPREP reads the data file (unit 31) which is set up with integer variables followed by real variables. This file is read in free format, so spaces are required between data fields. The program can be modified to handle either formatted reads or binary reads by modifying the

appropriate sections of the program (two open statements and two read statements). The data file is read twice, first to determine the number of levels for each discrete factor and the simple statistics for each continuous variable (covariates and traits), and second to recode levels of factors to correspond to the equation numbers in the MME and to express each continuous variable as a deviation from its mean. After the second read of the data the equations which are to be blocked with animal effects are written to file MTGS44. An equation is blocked with animal effects only if that equation is coded in the same column as the animal effect or a second animal effect for a trait.

To illustrate the strategy used in the programs, data from Meyer (1991) will be used. Records are body weight (t1) and intake (t2) measured on 284 animals. The pedigree file includes 45 base animals for a total of 329 animals in **A**. For each trait the effects in the model of analysis include random animal (a), maternal genetic (m), and maternal permanent environment (pe: 42 levels). Fixed effects are litter size (lsc: covariate) and generations (gen: 3 levels) for body weight, and litter size (lsd: 7 levels) and sex (sex: 2 levels) for intake. The first two records in the data file (7 integers and 3 reals) are:

animal	sire	dam	gen	sex	lsd	lit	lsc	t1	t2
20101	11012	10101	1	1	4	1	4.0	22.5	59.1
20102	11012	10101	1	1	4	1	4.0	22.6	0.0

Note that litter size appears twice in the data, both as the sixth integer (lsd) and as the first real (lsc) variable. In addition, intake was deleted for animal 20102 to demonstrate the effect of missing data so the field for t2 is coded as 0.0 which is used as the missing value.

The first step is to run MTGSNRM which forms $\mathbf{A}^{-1}$ and writes the sorted vector of 329 animal IDs to unit 21 (ascii) and the half stored non-zero elements to unit 22 (binary). MTGSPREP is then run with the following parameters in input file *example.in*; MTGSPREP is executed using the DOS command *mtgsprep.exe* < *example.in* (see page 13 for a description of using input files rather than entering input values interactively):

```
PREPEX.DAT                      name of data file
test of MTGSPREP with km mouse data
2 traits
*                               end of comments
7                               number of integers on each line of data file
3                               number of reals on each line of data file
2                               number of traits in analysis
weight                          name of trait 1
2                               position of trait 1 in vector of real values
0.0                             value of missing observation for trait 1
1                               number of covariates
litter size                     name covariate 1
1                               position of covariate 1 in vector of real values
1                               maximum power for covariate 1
1                               number of fixed factors
generation                      name for fixed factor 1
4                               position of fixed factor 1 in vector of integers
0                               write summary of fixed factor 1 levels to log file
1                               position of animal effect in vector of integers
329                             num. of animals in relationship matrix (from NRM)
1                               include second animal effect
mat gen                         name of second animal effect
3                               position of second animal effect in vector of integers
1                               number of uncorrelated random factors
mat pe                          name of uncorrelated random factor
3                               position of uncorr. rand. factor in vector of integers
0                               do not write summary of uncorr. ran. factor to log
intake                          name of trait 2
3                               position of trait 2 in vector of reals
0.0                             value for missing observation for trait 2
0                               number of covariates
2                               number of fixed factors
litter size                     name for fixed factor 1
6                               position of fixed factor 1 in vector of integers
0                               write summary of fixed factor 1 levels to log file
sex                             name of fixed factor 2
5                               position of fixed factor 2 in vector of integers
0                               write summary of fixed factor 2 levels to log file
1                               include second animal effect
mat gen                         name of second animal effect
3                               position of second animal effect in vector of integers
1                               number of uncorrelated random factors
mat pe                          name of uncorrelated random factor 1
```

3	position of uncorr. rand. factor in vector of integers
0	write summary of uncorr. ran. factor to log file
1	write labels for cov. and fixed factors to MTGS45
1	write labels for uncorr. random factors to MTGS46

The subsequent steps are:

1. Read the number of animal IDs from unit 31 and compare to the number entered by the user as read from the log of MTGSNRM (i.e., 329). If these numbers are not equal, the program terminates because the wrong pedigree file is probably being used. If the numbers are equal, the sorted vector of IDs is read from unit 21.

First read of data:

2. The 284 lines of data are then read sequentially from unit 31. For each line, all (7) integer variables are read into an integer vector and all (3) real variables are read into a real vector. Each of the j = 2 traits is then processed:
   a) If the value for the trait is equal to the missing value (0.0), skip to the next trait.
   b) If the value for the trait is valid:
      i) update the count, sum, and sum of squares for each real variable,
      ii) compare the class value of each fixed factor and uncorrelated random factor (e.g., pe) to the unique numeric (but unsorted) list of current values stored in memory; if the level is not already in the list, it is added at the end and the number of levels for the effect is incremented by one.
3. After all lines of data have been read into memory, sort the vectors of levels for each discrete fixed and uncorrelated random factor.
4. Calculate the mean and variance for each real variable.
5. Based on the sequence of the MME and the number of levels for each factor, determine the starting row of each factor in the model. The starting row is expressed as one less than the actual position. For the example data the structure is:

Factors	(starting row MME)-1	No. rows
t1: linear covariate for litter size	0	1
t1: fixed effect for generation	1	3
t2: fixed effect for litter size	4	7
t2: fixed effect for sex	11	2
t1: random animal	13	329
t2: random animal	342	329
t1: random second animal (mat gen)	671	329
t2: random second animal (mat gen)	1000	329
t1: random maternal pe	1329	42
t2: random maternal pe	1371	42

For this example, there are 1413 equations in the MME. The means for trait 1 are 4.48 for the litter size covariate and 24.07 for weight, and for trait 2 the mean for intake is 64.30.

Second read of data:

6. The 284 lines of data are then reread sequentially from unit 31. For each line, all (7) integer variables are read into an integer vector and all (3) real variables are read into a real vector. Each of the j = 2 traits is then processed:
   a) If the value for the trait is equal to the missing value (0.0), skip to the next trait.
   b) If the value for the trait is valid, the value and position of each element in **W** and **y** is then determined:
      i) covariates and observations are deviated from their corresponding means, e.g., the deviations for the first record are: 4.0-4.4789=-0.4789 (litter size), 22.5-24.0687=-1.5687 (weight), and 59.1-64.2975=-5.1975 (intake).
      ii) the **W** row position of each regression coefficient is determined from the sequence and order (linear, quadratic, etc.) of the covariates,
      iii) the position of each discrete factor in **W** is determined by looking up its position in the corresponding vector of sorted levels and then adding this position number to the starting row position for the factor; e.g., for the first record, the value of 4 for litter size

(trait 2) is found at position 4 in the sorted list of levels (1, 2,..., 6, 7) so 4 is added to the starting position (4) for litter size to give the row position in **W** of 8.

7. Let $\mathbf{W}_{ij}$ be the element of **W** from row i and column j. After each line of data is read, the values and positions of elements in $\mathbf{W}_{ij}$ and the value for $\mathbf{y}_{ij}$ are written to unit 42 (binary); the length of $\mathbf{W}_{ij}$ is determined by the total number of model effects and the number of valid traits, and the number of $\mathbf{W}_{ij}$ rows is equal to the number of valid traits.
8. After all lines of $\mathbf{W}_{ij}$ have been written to unit 42 for a record, the column positions are written to unit 42 and a summary for each animal is written to unit 43; this information consists of number of effects (rows in $\mathbf{W}_{ij}$), data lines, trait combination number (1 to 2^n-1) , and structure (i.e., pattern of missing values) of observations for the animal. For the record of animal one, the values written to unit 42 and 43 are (subscript denotes trait number; text in parentheses is not written):

unit 42 - values and positions for $\mathbf{W}_{ij}$:

	(lsc_1)	(gen_1)	(lsd_2)	(sex_2)	(a_1)	(a_2)	(m_1)	(m_2)	(pe_1)	(pe_2)	(y_i)
(t_1)	-0.479	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0	-1.5687
(t_2)	0.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	1.0	-5.1975
(row)	1	2	8	12	59	388	692	1021	1330	1372	

unit 43 - summary of information for first animal:

		(Trait structure)		
(No. effects)	(No traits)	(code)	(p_1)	(p_2)
10	2	3	1	1

Trait structure codes are used to indicate the form of $\mathbf{R}_i^{-1}$ to be used in $\mathbf{W}_i'\mathbf{R}_i^{-1}\mathbf{W}_i$ and $\mathbf{W}_i'\mathbf{R}_i^{-1}\mathbf{y}_i$, and consist of a value (p_j) for each trait of 1 or 0, if the trait is present or missing, respectively, and a code calculated as $\sum_{j=1}^{t}\left(p_j \times 2\right)^{(j-1)}$. In the record of the second animal, trait 2 (intake) is missing so a single $\mathbf{W}_{ij}$ row of five effects is written to unit 42:

	(lsc_1)	(gen_1)	(a_1)	(m_1)	(pe_1)	(y_1)
(t_1:)	-0.4789	1.0	1.0	1.0	1.0	-1.4687
(row)	1	2	60	692	1330	

The information written to unit 43 is:

			(Trait structure)	
(No. effects)	(No. traits)	(code)	(p_1)	(p_2)
5	1	1	1	0

9. After the data have been read the second time, the blocking information is written to unit 44. The equations to be blocked with genetic effects are determined by finding uncorrelated random effects coded in the same column of the original data set as the animal effects. In addition, equations may be blocked with genetic effects if a trait has a second animal effect and there is an uncorrelated random effect coded in the same data column as the second animal effect for that trait. Note that the second animal effect can be different for traits (i.e., maternal genetic effect or paternal genetic effects can be fit for different traits). The form of the data written to unit 44 is animal number (renumbered) and the number of blocked equations on the first line, and the equation numbers to be blocked for that animal on the following line. The information for the first animal in the blocking file is:

(animal)	(No. Equations)
21	2
(equations)	
1330	1372

Animal 21 (original ID 10101) does not have a record, although it is present as a dam for the first group of animals in the data set. As a result, the maternal permanent environmental effect equations for that animal are blocked with the genetic effects because it is coded in the same column as the maternal genetic effect for both traits. Notice that the maternal permanent environmental equations are blocked with the DAM not with the animal having the record. This is because the dependency in the model occurs between maternal genetic and maternal permanent environmental effects (see the discussion of blocking in Chapter 3 on page 56).

10. Finally, information describing the models and MME (e.g., number of traits, starting position and number of levels for each effect) is written to unit 41 (ascii). If requested, the labels for

covariates and fixed effects are written to unit 45 and labels for uncorrelated random effects are written to unit 46. A summary of the model and data is then written to unit 82 (MTGS82 in ascii).

MTGSRUN

Linked List Matrix Storage

Several unique strategies were employed in implementing the Gibbs sampler. First, the MME are built in two pieces; the first part includes the least squares part of the equations (i.e., $\mathbf{W}'\mathbf{R}^{-1}\mathbf{W}$) with the $\boldsymbol{\Sigma}$ augmented only for the uncorrelated random effects. The $\mathbf{G}^{-1}\otimes\mathbf{A}^{-1}$ is not explicitly added to this matrix. Instead, one copy of $\mathbf{A}^{-1}$ is stored and used in the iterative algorithm. Define:

$$\mathbf{C}^* = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^* \end{bmatrix}, \text{ and}$$

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \bigoplus_{i=1}^{\gamma} \left[\mathbf{D}_i^{-1} \otimes \mathbf{I}_{r_i} \right] \end{bmatrix}.$$

Sparse matrix techniques are used to store only the non-zero elements of $\mathbf{C}^*$. In addition, the non-zero elements are half-stored to further reduce memory requirements. The non-zero elements are stored in a linked list form (as described by George and Liu, 1980). The subroutines used to build the linked list are based on those distributed with the set of ITPACK programs, which is a set of programs used to solve sparse linear systems of equations using iteration. The universal resource locator (URL) for world wide web (WWW) browsers for information about ITPACK and the subroutines is *http://www.netlib.org/itpack/index.html* . The subroutines are used to accumulate the non-zero elements by row and column and then convert the information to the linked list form. The linked list consists of 3 vectors, for example, IA, JA, and A, where the vectors are of length at least nr+1, nz, and nz, and declared as integer, integer, and double precision, respectively, and nr is the number of rows (and columns) in the matrix, and nz is the number of non-zero elements. Element i of IA is a pointer to the starting location of the elements in row i. The last element of IA is nz+1. The elements of JA contain the column numbers for the non-zero elements, and A contains the non-zero elements of the matrix. That is,

the elements of row i are stored in elements IA(i) to IA(i+1)-1 of A, and the column identifiers in the same elements of JA.

To illustrate the format, consider the linked list storage of the upper half-stored elements of

$$\mathbf{AA} = \begin{bmatrix} 3 & 0 & 1 & 0 \\ 0 & 5 & 0 & 2 \\ 1 & 0 & 3 & 1 \\ 0 & 2 & 1 & 1 \end{bmatrix}, \text{ then } \mathbf{IA} = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 8 \end{bmatrix} \mathbf{JA} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \\ 3 \\ 4 \\ 4 \end{bmatrix} \mathbf{A} = \begin{bmatrix} 3 \\ 1 \\ 5 \\ 2 \\ 3 \\ 1 \\ 1 \end{bmatrix}.$$

From the linked list, the dense half-stored matrix can generated from this list using the following Fortran statements:

```
DO I = 1,NR
        DO J = IA(I),IA(I+1)-1
                DENSEA(I,JA(J))=A(J)
        END DO
END DO
```

An additional subroutine was written to rebuild the MME without rebuilding the linked list structure. To do this non-zero elements are sorted by columns within a row in two groups after the linked list is built. The list is sorted within the blocked equations and within the unblocked equations. This allows the linked list structure to be searched to find the location of a non-zero element and update the value. Note that because the structure is recycled zero elements must be stored in locations where non-zero elements may be generated with different variance matrices. For example, if the linked list is built using a starting covariance between uncorrelated random effects of zero, the location where non-zero elements could result from non-zero covariances must be included to allow reuse of the linked list.

Finally, due to reasons to be outlined when discussing the Gauss-Seidel and Gibbs sampling implementations, there are some elements of the coefficient matrix which are upper half-stored, and some lower half-stored. Some elements must be lower half-stored for some of the blocked equations so that the blocked algorithms function properly.

In addition to **C*** being sparse stored, $\mathbf{A}^{-1}$ is stored using a linked list. This matrix, however, is **full** stored. Full storage is required for two reasons. The primary reason is that the implementation of the Gauss-Seidel and Gibbs sampling algorithms require access to a full row or column of $\mathbf{A}^{-1}$, which is not possible when the elements are half stored in a linked list. Calculation of quadratics of the form $\mathbf{u}'\mathbf{A}^{-1}\mathbf{u}$ is also easier with $\mathbf{A}^{-1}$ full stored, although it is possible with the matrix half stored.

Gauss-Seidel Iteration

Gauss-Seidel iteration is available as an option in the MTGSAM programs to allow solutions to be obtained for the MME for a given set of (co)variances. Define $s_i^{(\kappa)}$ as the update to solution i in round κ of iteration and neqn as the number of equations, then the update algorithm for Gauss-Seidel iteration can be written as (Golub and Van Loan, 1989):

initialize **r**, the right hand side
for i = 1 to neqn,

$$s_i^{(\kappa+1)} = \frac{1}{c_{ii}}\left(r_i - \sum_{j=1}^{i-1} c_{ij} s_j^{(\kappa+1)} - \sum_{j=i+1}^{neqn} c_{ij} s_j^{(\kappa)} \right)$$

end. [4.1]

This form can be rewritten for a blocked algorithm by replacing the scalar elements with blocks of coefficients. The Gauss-Seidel iteration algorithm can be considered intuitively as adjusting the right hand sides (RHS) of the equations for the current value of all the other effects in the model (i.e., the off-diagonal elements of the coefficient matrix) and scaling by the inverse of the diagonal element of the coefficient matrix.

In the case where the equations are upper half stored [4.1] can be rewritten as:

initialize **r*** as **r**, the RHS
for i = 1 to neqn,

$$s_i^{(\kappa+1)} = \frac{1}{c_{ii}}\left(r_i^* - \sum_{j=i+1}^{neqn} c_{ij} s_j^{(\kappa)} \right)$$

for j = i + 1 to neqn

$$r_j^* = r_j^* - c_{ij} s_i^{(\kappa+1)}$$

end
end. [4.2]

An analogous blocked form of this algorithm also exists. This algorithm accesses the elements of the coefficient matrix in an efficient sequence when the elements are upper half stored.

Finally, a hybrid of the two forms of the algorithm can be generated using the set of equations where the coeffient matrix is not augmented with $\mathbf{G}^{-1}\otimes\mathbf{A}^{-1}$ (i.e., $\mathbf{C}^*$) and using $\mathbf{G}^{-1}$ and $\mathbf{A}^{-1}$. First define $\mathbf{a}^{-i}$ as row i of $\mathbf{A}^{-1}$ with the diagonal element, a^{ii}, removed, $\mathbf{s}_{j,-i}^{(\kappa)}$ is the n-1×1 vector of genetic effects from round κ of iteration for trait j with element i removed, and $g^{i,j}$ is element i,j of $\mathbf{G}^{-1}$. Let $\mathbf{d}_i$ be the vector of adjusted RHS for the block of equations for animal i, and let $\mathbf{C}_i$ be the matrix of elements of the coefficient matrix associated with the blocked equations which is formed from $\mathbf{C}^*$, $\mathbf{G}^{-1}$, and a^{ii}. Finally, let $p_{j,i}^{(\kappa)} = \mathbf{a}^{-i}\mathbf{s}_{j,-i}^{(\kappa)}$. Using the special form of the matrix where $\mathbf{C}$ is written as the sum of $\mathbf{C}^*$ and $\mathbf{G}^{-1}\otimes\mathbf{A}^{-1}$, equations [4.1] and [4.2] can be rewritten as:

initialize $\mathbf{r}^*$ as $\mathbf{r}$, the RHS
update fixed effects and $\mathbf{r}^*$ using [4.2]
for block of equations for animal i, i = 1 to n,
  if equation j in the block corresponds to an animal effect for trait ℓ update the RHS as $\mathbf{r}_j^* = \mathbf{r}_j^* - \sum_{k=j+1}^{neqn} \mathbf{c}_{jk}^* s_j^{(\kappa)} - \sum_{k=1}^{t} g^{\ell,k} p_{k,i}^{(\kappa)}$
  if equation j is in the block corresponds to an a blocked uncorrelated random effect update the RHS as $\mathbf{r}_j^* = \mathbf{r}_j^* - \sum_{k=j+1}^{neqn} \mathbf{c}_{jk}^* s_j^{(\kappa)}$
  update the block of genetic and uncorrelated random effects as $\mathbf{C}_i^{-1}\mathbf{d}_i$
  update RHS for equations following block (k) for all solutions in block (j): $r_k^* = r_k^* - c_{jk} s_j^{(\kappa+1)}$
end
update unblocked uncorrelated random effects and $\mathbf{r}^*$ using [4.2] [4.3]

Fixed and Unblocked Uncorrelated Random Effects

The scalar implementation of [4.2] is used for these effects. The linked list is built in a specific form to optimize the Gauss-Seidel iteration algorithm for these equations: the diagonal element is stored in the first location in the group of elements for a row in the linked list followed by a sorted list of the upper half stored off diagonal elements in the row.

Genetic and Blocked Uncorrelated Random Effects

There is a subtle potential problem when implementing the blocked version of [4.3] when uncorrelated random effects are included in the block. The problem can best be demonstrated by example. Consider the following example coefficient matrix with 4 equations where the

numbers indicate upper half stored non-zero elements and an x represents the symmetric element not stored:

$$\begin{bmatrix} 1 & & & 2 \\ & 3 & 4 & \\ & x & 5 & 6 \\ x & & x & 7 \end{bmatrix}.$$

If equations 1 & 4 and 2 & 3 are treated as blocks, then for the evaluation of the first block (equations 1 and 4) the diagonal matrix, $\mathbf{c}_{ii} = \begin{bmatrix} 1 & 2 \\ x & 7 \end{bmatrix}$, is determined correctly. However, when

the adjustment for the off-diagonal elements is considered, the off-diagonal elements between equations 3 and 4 will not be found if the equations are upper half stored. Another way to visualize the problem is to consider the form of the coefficient matrix where elements from equations 1 & 4 are permuted to equations 1 & 2 and coefficients from equations 2 & 3 are permuted to equations 3 & 4. The resulting coefficient matrix is:

$$\begin{bmatrix} 1 & 2 & & \\ x & 7 & & x \\ & & 3 & 4 \\ & 6 & x & 5 \end{bmatrix}.$$

In this form, it is easier to see the problem: the matix once permuted is no longer upper half stored. To account for this problem some coefficients are upper half stored and some lower half stored when building the linked list. Coefficients for fixed effects and for unblocked uncorrelated random effects are upper half stored. For coefficients for genetic and blocked uncorrelated random effects, the animal associated with the equation is determined so that the order of the updating is known. Elements are upper half stored if the animal associated with the row is less than or equal to the animal associated with the column, and lower half stored otherwise.

Checkpointing

Checkpointing is the practice of saving intermediate results in an analysis so that a program can be continued from some point without having to restart the program and losing computer time. The MTGSRUN program checkpoints in two ways. MTGSRUN saves information to unit 90 entered by the user that will not change through the analysis. This information includes comments or description entered, prior distribution information, grouping information for uncorrelated and residual effects, Gauss-Siedel convergence criteria and maximum iterations, Gibbs sampling iteration information including total rounds, length of burn-in, and checkpoint frequency, and contrast coefficients. The program saves information alternately to units 91 and 92 at intervals specified by the user. The information written to these files include means for variance components, heritabilities, and correlations, inverses of (co)variance matrices, and the random number generator common block information. After the information is written to the checkpoint file, that file is closed and the files containing the samples (unit 61) and parameters (unit 62) from the Gibbs chain are closed and reopened. The files must be closed to be certain that the information is actually written to the file by the operating system, otherwise the system could lose the data in a buffer in a system crash. In the case of a restart from the checkpoint information, the programs will determine which of the two changing checkpoint files was written from the later round of iteration and attempt to open that file. If there is an error in the first checkpoint file tried the program will attempt to use the second file; there should be little chance of both files being corrupt.

Gibbs Sampling

As shown in Chapter 3, the full conditional (i.e., sampling) distributions for the fixed and random effects are all normal distributions. Therefore, only means and (co)variances are needed to generate these effects. The calculation of mean of the distribution is identical to the updating algorithm for the Gauss-Siedel iteration algorithm described earlier, and therefore, the algorithms used are very similar. The Gibbs sampling algorithm is simply an extension of the Gauss-Seidel iteration algorithm; once the mean is determined using the Gauss-Seidel iteration algorithm a normal deviation is added to that mean based on the (co)variances of the full conditional distribution. In the scalar case, which is used for fixed effects and unblocked uncorrelated

random effects, the variance is equal to the inverse of the diagonal element of the coefficient matrix. For the blocked algorithm, which is used for genetic effects and blocked uncorrelated random effects, the variance is the inverse of the matrix of elements of the coefficient matrix corresponding to equations in the block (a block diagonal matrix of the permuted coefficient matrix). The deviations from the mean are calculated as **Lr**, where **L** is the Cholesky decomposition of the (co)variance matrix **B** and **r** is a vector of uncorrelated standard normal (mean = 0, SD = 1) deviates. The Cholesky decomposition of **B** is a matix **L**, such that **B** = **LL**'. Therefore, VAR(**Lr**) = **L**VAR(**r**)**L**' = **LIL**' = **B**. In the scalar case, the problem simplifies because **L** is simply the square root of **B** and **r** is a single standard normal deviate.

The generation of the variance components is straightforward. The quadratic forms are calculated, based on the form of the full conditional distribution derived in Chapter 3, the prior distribution information is added (if a non-flat prior is used), and a new (co)variance matrix is generated.

Random Number Generation

All random number generation subroutines were written in Fortran. The random number generator used to generate standard (0,1) uniform random variables was based on one described by Marsaglia and Zamin (1987). The random number generator for normal random variables is based on the Kinderman and Ramage (1976) procedure as described by Kennedy and Gentle (1980). The random number generator for Wishart random variables is based on the algorithm described by Odell and Feiveson (1966) and discussed by Kennedy and Gentle (1980). A random number generator for Chi-square random variables was needed for that algorithm, and an approximation based on Wilson and Hilferty (1931) was used.

CHAPTER FIVE: Output File Formats

Because (as of this writing) the MTGSAM programs generate only the Gibbs sampling chains and calculate mean estimates, it is important for the user to be able to access the Gibbs sampling information generated to evaluate burn-in, to evaluate the thinning rate (i.e., the sampling lag for independent samples), and most importantly, to estimate posterior distributions. There are three files are needed to access the samples, units 61, 62, and 63 (i.e., MTGS61, MTGS62, and MTGS63). Files MTGS61 and MTGS62 are written in direct access unformatted format. The unit 61 file contains the observed values of the variance components, contrasts (if any), and solutions (if requested). The unit 62 file contains the parameters used to generate the samples from the appropriate distribution. The parameters needed are shape and scale parameters for inverted Wishart variables and mean and variance for normal variables. Because the shape parameter for the the inverted Wishart variables does not change, it is not written with each sample; it is written once to MTGS63. For contrasts, the mean and variance is only known in the single variable situation. That is, for contrasts containing more than one element, the closed form of the distribution is not known. Therefore, the unit 62 file has the mean and variance only for single element contrasts. For multiple element contrasts, the observed value is written in place of the mean of the distribution and a value of 0 is written in place of the variance. The order of the variables written to unit 62 for the contrasts and solutions is $mean_1$, $variance_1$, $mean_2$, $variance_2$, etc.

MTGS63 contains information needed to to read units 61 and 62. The mouse data used to describe the MTGSPREP program will be used as an example. Assume that MTGSPREP has been run using the input file from page 68 and MTGSRUN has been run using the following input file (note that the blank lines are required):

```
1                                   Type of run - New analysis
Mouse data from Karin Meyer
Multiple trait analysis of
Body weight and feed intake
*

1       2.47783750007415800         Genetic (co)variance means
```

2	.459944782089452000	Genetic (co)variance means
3	1.57927675355541000	Genetic (co)variance means
4	-.594199892718565200	Genetic (co)variance means
5	8.19549233549404500	Genetic (co)variance means
6	1.02224752348769800	Genetic (co)variance means
7	.121688323984633000	Genetic (co)variance means
8	1.10564369918392600	Genetic (co)variance means
9	-.370200664457456500	Genetic (co)variance means
10	.162979687500515000	Genetic (co)variance means
0	.000000000000000000	Done entering values
1		Genetic priors are correct
9		Genetic (co)variance priors shape parameter
0		No uncorrelated random covariances restricted
1	.827308600070143100	Uncorrelated random (co)variance means
2	-1.69664802413357700	Uncorrelated random (co)variance means
3	3.51285576291655400	Uncorrelated random (co)variance means
0	.000000000000000000	Done entering values
1		Ind Rand priors are correct
9		Ind Rand (co)variance priors shape parameter
0		No residual covariances restricted
1	5.34178427718693300	Residual (co)variance priors
2	3.80577188391943900	Residual (co)variance priors
3	12.3974357127453400	Residual (co)variance priors
0	.000000000000000000	Done entering values
1		Residual priors are correct
9		Residual (co)variance priors shape parameter
200		Rounds of Gauss-Seidel iteration
.100000000000000000E-04		Convergence criteria for Gauss-Seidel
10000		Rounds of Gibbs sampling (includes burn-in)
2000		Rounds of burn-in before writing Gibbs sampling
1		Frequency of writing Gibbs samples
100		Frequency of check-pointing
1		Write out all solutions? (Y=1,N=0)
1		Write out user specified contrasts
2		Number of contrast to monitor
2		Number of coeficients
12	1.00000000000000000	Contrast equation and coefficient
13	-1.00000000000000000	Contrast equation and coefficient
1		Number of coeficients
14	1.00000000000000000	Contrast equation and coefficient
6210	19906	Seeds for random number generator

Then, the contents of MTGS63 are:

STRT61	STRT62	LENGTH	SHAPE	#	DESCRIPTION
1	1	10	338		GENETIC (CO)VARIANCES
11	11	3	51	1	INDEPENDENT RANDOM (CO)VARIANCES
14	14	3	293	1	RESIDUAL (CO)VARIANCES
17		2			OBS CONTRASTS TO IUN61
	17	4			MEAN AND VAR OF CONTRASTS TO IUN62
19		1413			OBS 'SOLUTIONS' IUN61
	21	2826			MEAN AND VAR OF 'SOLUTIONS' TO IUN62

Total number of equations: 1413

Recall that both MTGS61 and MTGS62 are written in binary form using direct file access. All variables are written using a double precision (real*8) format. The first column in MTGS63 contains the starting location for each of the variables written to unit 61. The second column contains the analogous starting location for unit 62. The next column contains the number of variables written to the appropriate file. Column four contains the shape parameters for the inverted Wishart variables, and column five contains the group number for uncorrelated random and residual variances. The group number corresponds to the information on grouping entered by the user. The final column contains a brief description of the variables written.

Then, the order of variables in MTGS61 is:

observed genetic (co)variance element 1
observed genetic (co)variance element 2
⋮
observed genetic (co)variance element 10
observed uncorrelated random (co)variance element 1
observed uncorrelated random (co)variance element 2
observed uncorrelated random (co)variance element 3
observed residual (co)variance element 1
observed residual (co)variance element 2
observed residual (co)variance element 3
observed contrast 1
observed contrast 2
observed solution 1
observed solution 2
⋮
observed solution 1413.

The order of variables in MTGS62 is:

scale parameter genetic (co)variance element 1
scale parameter genetic (co)variance element 2
⋮
scale parameter genetic (co)variance element 10
scale parameter uncorrelated random (co)variance element 1
scale parameter uncorrelated random (co)variance element 2
scale parameter uncorrelated random (co)variance element 3
scale parameter residual (co)variance element 1
scale parameter residual (co)variance element 2
scale parameter residual (co)variance element 3
mean of contrast 1
0.D0 (since multiple element contrast)
mean of contrast 2
variance of contrast 2
mean of solution 1
variance of solution 1
mean of solution 2
variance of solution 2
⋮
mean of solution 1413
variance of solution 1413.

References

Casella, G. and E. I. George. 1992. Explaining the Gibbs sampler. Amer. Stat. 46:167.

George, A., and J. W. H. Liu. 1980. Computer solution of large sparse positive definite systems. Prentic-Hall, Englewood Cliffs, NJ.

Gianola, D. and R. L. Fernando. 1986. Bayesian methods in animal breeding theory. J. Anim. Sci. 63:217-244

Gianola, D., S. Im, and F. W. Macedo. 1990. A frameworkd for prediction of breeding value. *In* Gianola, D. and K. Hammond (Eds.) Advances in statistical methods for genetic improvement of livestock. Springer-Verlag, Berlin, pp 210-238.

Henderson, C. R. 1950. Estimation of genetic parameters. Ann. Math. Stat. 21:309.

Henderson, C. R. 1963. Selection index and expected genetic advance. In: Statistical Genetics in Plant Breeding. NAS-NRC publication 982.

Henderson, C. R. 1975. Best linear unbiased prediction under a selection model. Biometrics 31:423.

Henderson, C. R. 1984. Application of linear models in animal breeding. U. Guelph, Guelph, ON, Canada.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. N. von Krosigk. 1959. Estimation of environmental and genetic trends from records subject to culling. Biometrics 15:192.

Hobert, J. P. 1994. Occurences and consequences of nonpositive Markov chains in Gibbs sampling. Ph. D. Thesis. Cornell University. Ithaca, NY.

Jensen, J., C. S. Wang, D. A. Sorensen, and D. Gianola. 1994. Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. Acta Agric. Scand. 44:193-201.

Johnson, N. L. and S. Kotz. 1972. Distributions in statistics. John Wiley & Sons, New York, NY

Kennedy, W. J., and J. E. Gentle. 1980. Statistical Computing. Marcel Dekker. New York, NY.

Kinderman, A. J., and J. G. Ramage. 1976. Computer generation of normal random variables. J. Amer. Stat. Assoc. 71:893.

Liu, J. S., W. H. Wong, and A. Kong. 1994. Covaraince structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. Biometrika. 81:27-40.

Marsaglia, G. and Zaman, A. 1987. Toward a more universal random number generator. Technical Report FSU-SCRI-87-50. Florida State University. Tallahassee, FL.

Odell, P. L. and A. H. Feiveson. 1966. A numerical procedure to generate a sample covariance matrix. J. Amer. Stat. Assoc. 61:199.

Scott, D. W. 1992. Multivariate Density Estimation: Theory, Practice, and Visualisation. John Wiley & Sons, New York, NY.

Van Tassell, C. P., G. Casella, and E. J. Pollak. 1994. Comparison of efficiency for blocked and scalar Gibbs sampling algorithms. J. Anim. Sci. 72(Suppl. 1):58(Abstr.).

Wang, C. S., J. J. Rutledge, and D. Gianola. 1994. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. Genet. Sel. Evol. 26:91-115.

Wilson, E. B., and M. M. Hilferty. 1931. The distribution of the Chi-square. Proc. Nat. Acad. Sci. 17:684.