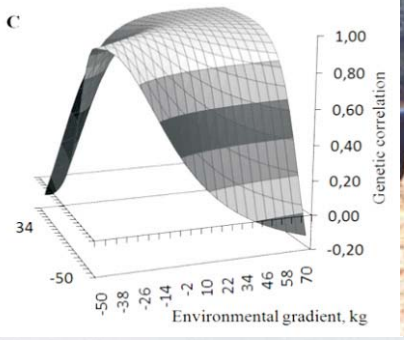
Documentos

ISSN 1982-5390
Dezembro, 2010
112

Application of Bayesian Inference in Animal Breeding Using the Intergen Program

Manual of Version 1.2

$$P = G + E + G \times E$$



Embrapa

ISSN 1982-5390
Dezembro, 2010

Empresa Brasileira de Pesquisa Agropecuária
Embrapa Pecuária Sul
Ministério da Agricultura, Pecuária e Abastecimento

Documentos 112

Application of Bayesian Inference in Animal Breeding Using the Intergen Program

Manual of Version 1.2

Fernando Flores Cardoso

Embrapa Pecuária Sul
Bagé, RS
2010

Exemplares desta publicação podem ser adquiridos na:

Embrapa Pecuária Sul
BR 153, km 603, Caixa Postal 242
96.401-970 - Bagé - RS
Fone/Fax: 55 53 3240-4650
http://www.cppsul.embrapa.br
sac@cppsul.embrapa.br

Supervisor editorial: Comitê Local de Publicações
Revisor de texto: Comitê Local de Publicações
Normalização bibliográfica: Graciela Olivella Oliveira
Tratamento de ilustrações: Roberto Cimirro Alves
Editoração eletrônica: Roberto Cimirro Alves
Foto(s) da capa e gráfico: Fernando Flores Cardoso

1ª edição online

Dados internacionais de Catalogação na Publicação (CIP)
Embrapa Pecuária Sul

Cardoso, Fernando Flores
Application of bayesian inference in animal breeding using o Intergen program : manual of version 1.2 [electronic resource] / Fernando Flores Cardoso. -- Bagé: South Animal Husbandry & Sheep, 2010.
(Documentos / South Animal Husbandry & Sheep, ISSN 1982-5390 ; 112)

System requirements: Adobe Acrobat Reader
Access mode: <http://cppsul.embrapa.br/unidade/publicacoes:list/259>
Web Page title (Accessed on 30 Dez. 2010)

1. Statistical methods. 2. Statistical analysis. 3. Animal breeding. I. Título. II. Série.

CDD 636.0821

Author

Fernando Flores Cardoso
Veterinarian, MSc., Ph.D,
Embrapa South Animal Husbandry &
Sheep Researcher,
Postal box 242, BR 153 Km 603,
CEP 96401-970 - Bagé, RS – Brazil
fcardoso@cppsul.embrapa.br

Presentation

Embrapa South Animal Husbandry and Sheep - CPPSUL is constantly developing new approaches and methods of scientific research on animal production systems.

In this context, Embrapa CPPSUL cattle and sheep breeding programme is focusing mainly on improvement of meat quality and animal adaptation to environmental conditions.

This paper presents the application of Bayesian inference in Animal Breeding Program using Intergen - Manual Version 1.2. This program aims to contribute to the animal breeding herds in the Campaign of Rio Grande do Sul and Brazil.

Thus, Embrapa offers this publication to scientists, technicians, consultants, students, farmer and others interested in improving economic caractheristics of cattle herds through Intergen software.

Roberto Silveira Collares
Chief General

Summary

Application of Bayesian Inference in Animal Breeding Using the Intergen Program
Manual of Version 1.2

Fernando Flores Cardoso

Introduction

The purpose of this document is to describe the functionalities and use of INTERGEN program – Version 1.2 originally described by Cardoso (2008). The software was developed at Embrapa Pecuária Sul in Fortran 90/95 language with capabilities to implement hierarchical Bayesian models and estimate their parameters using Markov Chain Monte Carlo methods (SORENSEN; GIANOLA, 2002).

INTERGEN is available on Embrapa Pecuária Sul's (EMBRAPA PECUÁRIA SUL, 2010). The possible models to be fit in INTERGEN include complex situations studied in quantitative genetics, using performance data from farm animals, including:

- Uncertainty paternity model, in the presence of multiple sires mating;
- Multibreed model, for crossbred populations with multiple breed compositions;
- Reaction norms model, in genotype-environment interaction studies;
- Robust model, for datasets with outliers;
- Structural variance model, to adjust the residual heterogeneity of variance.

The software is based on the BLUPF90 family of programs developed by Misztal et al. (2002) and the relies on libraries to generate random

numbers and to perform operations with sparse matrices (MISZTAL, 2009) to implement the mixed model (HENDERSON, 1984) and MCMC methodology.

The INTERGEN program, despite allowing complex modeling, has a very simple interface and is called from the DOS Command Prompt window or Linux terminal and is controlled by a parameter file, which contains information of data and pedigree files, the effects in the model and the MCMC chain to be implemented.

What's New in Version 1.2

A new Metropolis-Hastings algorithm was developed and implemented to allow fitting of reaction norms models with residual heterogeneity of variance. This allows for joint estimation of the residual variation through an exponential function of the environmental gradient and the values of the unknown environmental covariate, which is used to characterize the environmental gradient.

Parameter file

Once activated, the program prompts the user to specify a parameter file, which must be previously prepared in unformatted text (ASCII).

The parameter file presents a general format, with details described below. The sections are defined by headings in bold capital letters, which should appear exactly as described and within a single line of ASCII file (although in this manual they may appear in two lines due to line size limitation). In the subsequent line, the fields changed by the user are listed with optional fields within parentheses.

Comments can be included in the parameter file, starting the comment line with #.

Description title of the analysis in free format

MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
total number of MCMC cycles, burn in period, thinning interval to save samples

Example
400000 100000 100

SEED
integer

Details
Seed to start the random variables generation process, which allows repeating the same chain.

Example
123

RESTART: Y/N? [CYCLE_TO_RESTART]
y/n, (cycle for continue the chain)

Details
Reports if the analysis is continuation of chain already in progress? y = continue the previous chain (must inform the cycle to restart), n = start new.

Example
y 10001

DATAFILE NAME N_RECORDS
data file name, number of data lines

Details
Data files are unformatted text files (ASCII), containing data on all effects in the model in columns, with at least one blank space between columns, with one row for each record.
In the case of an animal model, it should contain the animal identification (ID) to

which the record pertains, in accordance to the coding used in the pedigree file. All classificatory effects must be recoded from 1 to the number of classes. There is no necessary order of effects in the file, however we recommend the following pattern: first column with animal ID; then the dam ID, in cases with maternal effect; then listing other classificatory effects, followed by the covariates; and the response variables. Missing data should be encoded with 0.

Example
manimal.dat 22124

NUMBER_OF_TRAITS
number of responses traits in analysis

Example
2

NUMBER_OF_EFFECTS
number of effects in the model

Example
5

OBSERVATION(S)
column(s) in the data file corresponding to the response variables

Details
Each column corresponds to one response variable.

Example
12 13 → two traits, with values in columns 12 and 13 of the data file.
10 → one trait, with values in column 10 of the data file.

WEIGHT(S)
Column(s) with the weights for the response variables

Details

One column per variable response; leave blank for analysis without weight.

Example

11 → weight on column 11 of data file.
→ without weighting.

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]

Position(s) in the data file - one column per response variable, number of levels, type of effect, save samples? **y**/**n**
(position(s) where the effect is nested - one column per variable response)

Details

EFFECTS in the model section with one row for each effect. The possible effect types in INTERGEN are:

- **cross** = classificatory,
- **cov** = covariate,
- **unknowncov** = classificatory effect to be used as covariate for reaction norms through random regression;
- **rnorm** = linear reaction norms through random regression;
- **ram** = reduced animal model effect with multiple sires.

The option *save samples*? **y=yes** should be used with caution, particularly for effects with many levels, to avoid the generation of extensive files, which takes a long time to be writing on hard disk.

The option [EFFECT NESTED] is used to specify random regression effects and reaction norms with unknown covariates, where:

→ *position in data file* = position of the covariate or orthogonal polynomial (one column per response variable);
→ *number of levels* = number of levels of nested effect (animal effect, for example);
→ *effect type* = **cov** for random regression model and **rnorm** for reaction norms model with unknown covariates;
→ *save samples?* (**y** = yes **n** = no);

→ *position where the effect is nested* = for example, position of animal effect in the data file (one column per response variable).

Example

4 1000 cov n 1 → polynomial values in column 4 of the data file, 1000 levels for the nested effect, **cov** type, not saving samples (the recommended option in this case), and nested effect (animal, for example) in column 1.

1 1520 cross n
2 1520 cross n
3 1 cov n
4 50 cross n → univariate model with four effects, three classificatory and one covariate.

Different models can be specified for each trait. The zero (0) value to an effect position in the data file is used to encode the absence of such effect for a particular trait.

Example (MISZTAL, 2007)

1 0 1000 cross n
0 7 2000 cross n → Variable 1 has contemporary group (CG) at position 1 with 1000 levels, and variable 2 has CG in position 7 with 2000 levels.
1 7 2000 cross n → same example in one line, where the number of levels is maximum between effects.

In the case of reduced animal model (effect type = **ram**), these effects must be the first in the list of EFFECTS section.

RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
residual distribuition type, prior degrees of belief

Details

There are six possible residual distribution assumption types, briefly, we have:
• **homogeneous** = Homoskedastic normal: Normal distribution (Gaussian) with

homogeneous variance $e \sim N(0,\sigma^2_e)$ – which is the usual assumption;

IMPORTANT: This is the only option for multivariate analysis. The following are valid only for univariate analysis.

- **strutural** = Heteroskedastic normal: Normal distribution with heterogeneous variance $e_{(i)} \sim N(0,\sigma^2_{e(i)})$, where $\sigma^2_{e(i)}$ is the variance of the errors $e_{(i)}$ of *i* subclasses combination defined by a structural model;

- **student_t** = Homoskedastic Student *t*: Student *t* distribution with homogeneous variance $e_{(i)} \sim N(0,\sigma^2_e/w_{(i)}) \rightarrow y_{(i)} \sim t_{(v)}(0,V^2_e)$ - robust model;

- **struct_student_t** = Heteroskedastic Student *t*: Student *t* distribution with heterogeneous variance $e_{(i)} \sim N(0,\sigma^2_e/w_{(i)}) \rightarrow y_{(i)} \sim t_{(v)}(0,V^2_{e(i)})$ - robust model;

- **slash** = Homoskedastic Slash: Slash distribution with homogeneous variance $e_{(i)} \sim N(0,\sigma^2_e/w_{(i)}) \rightarrow y_{(i)} \sim Slash_{(v)}(0,V^2_e)$ - robust model;

- **struct_slash** = Heteroskedastic Slash: Slash distribution with heterogeneous variance $e_{(i)} \sim N(0,\sigma^2_{e(i)}/w_{(i)}) \rightarrow y_{(i)} \sim Slash_{(v)}(0,V^2e_{(i)})$ - robust structural model.

For the degrees of belief on random effects (co)variance(s) prior values, the higher the specified value the larger the influence of a priori values in the analysis results. Specifying the number of variables in analysis is equivalent to an diffuse priori and therefore the inference is basically driven by data information.

Example

homogeneous 1 → homoskedastic normal residuals with 1 degree of belief in the residual variance prior value.

METROPOLIS_STEP_OF_STRUCTURAL_EFFECTS: ROUNDS_WITHIN_CYCLE TUNING_SKIP
(optional)

number of Metropolis steps, tuning interval

Details
This optional section is only required in cases of a structural effect in the residual, *structural, struct_student_t* or *struct_slash* types.
In this section is specified the Metropolis-Hastings (MH) algorithm steps number within each chain cycle and the interval of cycles to adjust the proposal distribution variance in the MH algorithm during the burn in period (advanced options of the structural model).

Example
5 10

NUMBER_OF_STRUCTURAL_EFFECTS (optional)
number of effects for the residual variance

Details
This optional section is only required in cases of structural effect in the residuals, *structural, struct_student_t* or *struct_slash* types.
This section defines how many model effects also affect the residual variance.

Example
4

STRUCTURAL_EFFECTS: LINE_FROM_EFFECTS_SECTION SAVE_SAMPLES? (optional)
EFFECTS section line, save samples? (**y**/**n**)

Details
This optional section is only required in cases of residual structural effects, *structural, struct_student_t* or *struct_slash* types, and must have one line for each effect.
The EFFECTS section line refers to the line number of order of this linear model effects section where the effect that will also affect the residual dispersion was specified.
You can save scalar factors samples using **y** (= yes).

Example

4 y → use the effect of EFFECTS: section fourth line and save samples.
6 y
7 y
9 n → use the effect of EFFECTS: section ninth line and not save samples

RESIDUAL_PRIOR_(CO)VARIANCES

a priori values for the residual covariance matrix

Example

200 10
10 100 → For a bivariate model with prior variances in the diagonal and covariances off diagonal.
200 0
0 100 → Covariance is assumed null and kept at zero (0) during the analysis.

RANDOM_GROUP (optional)

line(s) of random effect(s) in EFFECTS section:

Details

This optional section is required when there are random effects in the model. You can specify more than one line, in cases of correlated random effects, like direct and maternal effects. Correlated effects must necessarily be consecutive in EFFECTS section. Other sections described below are necessary to complement the description of random effects. The RANDOM_GROUP section and subsequent are repeated for each effect or random effects group.

Example

1 2 → effects on the first and second line of the EFFECTS section are random and correlated.

RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF (optional)

random type, prior degrees of belief

Details

This optional section necessarily accompanies the *RANDOM_GROUP* section.
The random effect type can be one of the following:

- **diagonal** = uncorrelated effect;
- **add_sire** = sire additive effect;
- **add_animal** = animal additive effect;
- **add_an_ms** = animal additive effect with multiple sires and genetic groups;
- **add_an_mb** = multibreed animal additive effect with genetic groups.

Example

add_animal 2 → animal model effect with 2 degrees of belief in the prior value.

PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS] (optional)

(pedigree file name, number of animals), (number of genetic groups), (number of breeds)

Details

This optional section necessarily accompanies the *RANDOM_GROUP* section. The different random effects types require different pedigree files. In general, the pedigree file must have one row for each animal with at least three columns.
All animals must be included in the pedigree file identified from 1 to the total number of animals in the file, including base animals with unknown parents.
Is not necessary that parents have smaller numbers than the progeny, except for reduced animal model (effect type = **ram** and random type = **add_an_ms**).
When a sire, dam or both are unknown they must be coded with 0. Other columns will be needed depending on the random effect associated with the pedigree in question, as described below:

- **diagonal ou diag_mb**:

Does not require pedigree file, because the levels are not correlated within variable. Ex: random contemporary groups and permanent environmental dam effect.

- **add_sire**:

Animal number, sire number, maternal grandsire number

- **add_animal**:

Animal number, sire number, dam number

• **add_an_ms**:
Animal number, dam number, sire number (or -1 if animal has uncertain paternity - multiple sires), 1/d, indicating that the animal has progeny in the file (1=yes, 0=no).
Here, *d*=proportion of genetic variance attributed to Mendelian sampling for the particular animal.
• **add_an_mb**:
Animal number, sire number, dam number, proportion of breed 1, ..., proportion of breed n.

Fixed genetic groups estimated jointly with the animals additive effects on **add_an_ms** and **add_an_mb** types are specified with greater numbers than the number of animals in pedigree file. For example, if the total of animals is 100 and we want to create two genetic groups, the animals are assigned to them by phantom parents with numbers 101 and 102, depending on the group they belong to.

Example
mbreed1.ped 100 2 2 → *mbreed1.ped* file has 100 animals with two genetic groups of a population composed of two breeds.
A possible line of this file would be:
23 101 101 0.625 0.375 → for an animal number 23, with unknown parents, but associated with genetic group 101, 5/8 of breed 1 and 3/8 of breed 2.

MULTIPLE_SIRES: MAX_N_FOR_MCMC [FILE: NAME & DIMENSION] [DIRICHLET_PRIORS] (optional)
maximum number of multiple sires (MS) for MCMC inference, (MS file name, file size), (Prior Dirichlet distribution hyperparameters)

Details
Optional section, accompanies *RANDOM_GROUP* section only if the random effect type = **add_an_ms** or **add_an_mb**, defining inference on the uncertainty paternity when they are MS.
The *maximum number of potential parents to make inference* defines the type of inference regarding paternity probability. You should use 0 for analysis without MS with the random types **add_an_ms** and **add_an_mb**; 1 to use the average

relationship matrix, which is also used when the MS group size exceeds the maximum number of possible parents for MCMC inference.
If the analysis includes animals with MS, *MS file name* and *number of lines* must be informed.
In this file, the possible sires for each individual associated with a MS group by -1 in the column of sire number in the pedigree file. The MS file has the following structure for each animal:
number of potential sires
identification of sire 1
a priori probability of sire 1
...
identification of sire n
a priori probability of sire n

For example, for an animal with 3 possible sires (4, 5 or 6), with the same prior paternity probabilities, we have:
3
4
0.3333333333
5
0.3333333333
6
0.3333333333

When INTERGEN identifies in pedigree file an animal with uncertainty paternity, it search on MS file the possible sires. Therefore, in this MS file, the possible sires of each animal must appear in the same order that animals with uncertain paternity are listed in the pedigree file, and the structure described above must be repeated for each animal. The identification of parents must also be consistent with the pedigree file.

Finally, in this section the value for hyperparameters of the Prior Dirichlet distribution of paternity probability defines the priors for paternity probabilities. If the value = 0, it does not use the Dirichlet distribution. If the value is greater than 0, it uses this value for all alpha parameters of Dirichlet distribution in all MS groups. If the value is less than 0, use values user defined alphas from the MS file.

Example

10 animodel.ms 252 -1 → Calculates posterior paternity probabilities for groups with up to 10 MS. For groups with more than 10 possible sires uses the average parentage. The animodel.ms MS file has 252 lines and the alphas values of the Dirichlet prior will be the probabilities specified in this file.

METROPOLIS_STEP_OF_MULTIBREED_(CO)VARIANCES: ROUNDS_WITHIN_CYCLE TUNING_SKIP (optional)

Metropolis steps number, tuning interval

Details

Optional section, accompanies *RANDOM_GROUP* section only if the random type effect is multiracial (RANDOM_TYPE = **add_an_mb** or **diag_mb**). Sets the number of the Metropolis-Hastings (MH) algorithm steps within each chain cycle and interval of cycles to adjust the proposals distribution variance in the MH algorithm during the burn in period (advanced options of multibreed model).

Example

2 20 → Runs two MH cycles within each MCMC chain cycle and adjusts the proposals distribution variance based on the average acceptance rate every 20 cycles during the first half of the burn in period.

(CO)VARIANCES (optional)

Prior values for the random effects covariance matrix

Details

Optional section, necessarily accompanies *RANDOM_GROUP* section. The structure of the (co)variance matrix (G) allows for multiple random effects and multiple traits. Considering an example with two effects (*a* and *m)* and two traits (1 and 2), we have the following symmetric matrix, where there are' diagonal blocks of each effect with variance and covariance between the different traits, and blocks off the diagonal with the covariance between different effects and/or traits:

$$G = \begin{bmatrix} \sigma^2_{a_1} & \sigma_{a_1a_2} & \sigma_{a_1m_1} & \sigma_{a_1m_2} \\ \sigma_{a_1a_2} & \sigma^2_{a_2} & \sigma_{a_2m_1} & \sigma_{a_2m_2} \\ \sigma_{a_1m_1} & \sigma_{a_2m_1} & \sigma^2_{m_1} & \sigma_{m_1m_2} \\ \sigma_{a_1m_2} & \sigma_{a_2m_2} & \sigma_{m_1m_2} & \sigma^2_{m_2} \end{bmatrix}$$

Although these matrices are symmetric, all their elements must be specified in the *(CO)VARIANCE* section.

Example

200	10	10	0
10	100	0	15
10	0	50	2
0	15	2	300

→ Note that $\sigma_{a1m1} = \sigma_{a2m1} = 0$, (the covariance is assumed null between *a* and *m* effects on different traits and kept at zero during the analysis).

How to run the program

To run the program the parameters, data, and pedigree files (and multiple sires, if this is the case) and the *intergen1.exe* executable must be in the same directory.

Open a DOS command prompt window (or Linux terminal), go to the directory where the files were placed and type: **intergen1.2** (or **./intergen1.2**) <enter>. Then the program will request the *parameterfile* name, through the question:

name of parameter file?

You should then type the *parameterfile* name <enter>.

This response can be included in a text file (for example, *run*), such that the *run* file (ASCII) has the following content:
parameterfile

So, the program can be run by:
C:\...patch...\intergen1 < run > out <enter>

In this case, the INTERGEN gets the *parameterfile* name from the *run* file and all the screen output is saved in an *out* file.

IMPORTANT: Many program crashes may be associated with errors in the number of model effects levels and in the specification of effects positions. Often these errors are associated with the word "hash" or "ACCESS VIOLATION".

How to continue an ongoing chain

In MCMC inference, there is often a need to lengthen a finished chain to increase information content or because a chain was interrupted by power failure or the need to turn off the computer. In these cases, the INTERGEN allows to continue the chain for extra cycles, without having to go through another burn in period or to discard the cycles of the interrupted chain. This is possible through the section RESTART: Y/N? [CYCLE_TO_RESTART].

To make this possible, each time that the output files are written, the program also writes the last sample of all parameters in the *solutions*, *varcomp*, *structural_r* and *robustness_w* files. To continue/restart the chain, the user must observe in the first row of one of these files which was the last recorded cycle, for example, in *varcomp*:

Variance components after: 44000 rounds and burn-in of 10000.

In this case, the last cycle was 44000 and the chain must be restarted at 44001, changing the corresponding section in the parameter file as follows:

RESTART: Y/N? [CYCLE_TO_RESTART]
y 44001

IMPORTANT: Furthermore, there is a very important not automatic step to be completed by the user. In the files *solutionsam*, *varcompsam* and *loglike_rnd* the samples are saved more often than in the files referred above. The user must open these files with a text editor, go to the end of file and delete all samples saved after the last cycle recorded in the *solutions*, *varcomp*, *structural_r* and *robustness_w* files (44000 in the example above), and then save this files with their original names, without any extension.

Output files

The following output files are generated by the INTERGEN program:

Solutions: Posterior mean and standard deviation of the model effects levels after burn in and the last sample to continue chain.

Solutionsam: Samples of the model effects levels of EFFECTS section where option SAVE_SAMPLES = **y**.

Varcomp: Posterior mean and standard deviation of the variance components after the burn in period and the last sample to restart the chain. The RANDOM_GROUP effects appears first and in the same order that they are specified in *parameterfile.* And at last, the residual (co)variance components are listed.

Varcompsam: Variance components samples and, if any, effects of heteroskedastic structural model. Variance components appear in the same order of *varcomp*. Because these matrices are symmetric, only elements of below diagonal are listed, as follows: v11, v21, v22, v13, v23, v33... The structural effects multiplicative factors for the residual variance appear after the residual variance and in the same order they are specified in *parameterfile*.

structural_r: Posterior mean and standard deviation of the structural

effects in the residual variance after burn in and the last sample to restart the chain, listed in the same order they are specified in *parameterfile*.

robustness_w: Posterior mean and standard deviation of the weighing variables of robust model after burn in and the last sample to restart the chain, for each record, in the same order that they appear in the data file.

Loglike_rnd: Log of the conditional data sampling distribution (first stage of the hierarquical model) for model choice criteria and adjustment measures.

Loglike_obs: Information for model choice criteria and adjustment measures.

Mh_dbeliefchg: Tuning of multiplying scale factor of the proposal distributions variance in the MH algorithm (if applicable).

Mh_acceptance: Acceptance rate of proposed values in the MH algorithm (if any).

Msiresppr: Posterior mean and standard deviation of paternity probabilities after the burn in period and the last sample to restart the chain (if applicable). This file also saves the proportion of times that each candidate was chosen as true sire in MCMC cycle.

Msirespprsam: Paternity probabilities samples for the possible parents of the first animal that appears with uncertainty paternity in pedigree file.

References

CARDOSO, F. F. **Manual de utilização do programa INTERGEN – versão 1.0 em estudos de genética quantitativa animal**. Bagé: Embrapa Pecuária Sul, 2008. 74 p. (Embrapa Pecuária Sul. Documentos, 74).

EMBRAPA PECUÁRIA SUL. **Software INTERGEN 1.0**. Available at: <http://www.cppsul.embrapa.br/unidade/servicos/intergen>. Accessed on: 30 dez. 2010.

HENDERSON, C. R. **Aplications of linear models in animal breeding**. Guelph: University of Guelph, 1984. 462 p.

MISZTAL, I. **BLUPF90 family of programs**. Available at: <http://nce.ads.uga.edu/~ignacy/newprograms.html>. Accessed on: mar. 2007.

MISZTAL, I. **Ignacy Misztal**. Available at: <http://nce.ads.uga.edu/~ignacy/>. Accessed on: 30 dez. 2010.

MISZTAL, I.; TSURUTA, S.; STRABEL, T.; AUVRAY, B.; DRUET, T.; LEE, D. H. BLUPF90 and related programs (BGF90). In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 7, 2002, Montpellier. **Proceedings**... Montpellier: INRA, 2002. 1 CD-ROM. Communication no 28-07.

SORENSEN, D. A.; GIANOLA, D. **Likelihood, Bayesian and MCMC methods in quantitative genetics**. New York: Springer-Verlag, 2002. 740 p.

Appendix

In the following, we present a series of *parameterfiles* to run some frequently used quantitative genetic analysis models based on Bayesian methods. Examples including data, pedigree and parameter files are available at the Embrapa Pecuária Sul website (EMBRAPA PECUÁRIA SUL, 2010).

A.1. Animal model with proper priors

```
#Example 1 - Animal model
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
11000   1000   10
SEED
2
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
animodel.dat   240
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
2
OBSERVATION(S)
4
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]
1   300   cross n
3   2   cross n
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous 1
RESIDUAL_PRIOR_(CO)VARIANCES
150
RANDOM_GROUP
1
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_animal 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
animodel.ped   300   0
(CO)VARIANCES
100
```

A.2. Animal model using the Henderson's average numerator relationship matrix.

```
# Example 2 - Animal model-average numerator relationship matrix
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
11000   1000   10
SEED
2
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
animodel.dat 240
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
2
OBSERVATION(S)
4
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]
1   300   cross n
3   2   cross n
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous   10000
RESIDUAL_PRIOR_(CO)VARIANCES
150
RANDOM_GROUP
1
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_an_ms 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
animodelrm.peda   300   0
MULTIPLE_SIRES: MAX_N_FOR_MCMC [FILE: NAME & DIMENSION]
[DIRICHLET_PRIORS]
1   animodelrm.ms   252
(CO)VARIANCES
100
```

A.3. Animal model with Westell's genetic groups.

```
# Example 3 - Animal model with Westell's genetic groups MCMC_CHAIN:
TOTAL_CYCLES BURN_IN THINNING_INTERVAL
11000   1000   10
SEED
2
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
animodel.dat 240
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
2
OBSERVATION(S)
4
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]
1   301   cross n
3   2   cross n
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous   10000
RESIDUAL_PRIOR_(CO)VARIANCES
150
RANDOM_GROUP
1
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_an_ms 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
animodelrm.pedg   300   1
MULTIPLE_SIRES: MAX_N_FOR_MCMC [FILE: NAME & DIMENSION]
[DIRICHLET_PRIORS]
0
(CO)VARIANCES
100
```

A.4. Uncertainty of paternity model for multiple sires mating progeny.

```
# Example 4 - Uncertainty of paternity model
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
11000   1000   10
SEED
2
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
animodel.dat   240
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
2
OBSERVATION(S)
4
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]
1   300   cross n
3   2   cross n
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous   10000
RESIDUAL_PRIOR_(CO)VARIANCES
150
RANDOM_GROUP
1
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_an_ms 1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
animodelrm.ped   300   0
MULTIPLE_SIRES: MAX_N_FOR_MCMC [FILE: NAME & DIMENSION]
[DIRICHLET_PRIORS]
10   animodelrm.ms   252   -1
(CO)VARIANCES
100
```

A.5. Structural model for residuals variances, with variance heterogeneity and robustness to outliers.

```
# Example 5 - Structural model for residuals variances
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
11000   1000   10
SEED
2
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
animodelh.dat   240
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
2
OBSERVATION(S)
4
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT SAVE_SAMPLES? [EFFECT NESTED]
1   300   cross n
3   2   cross n
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
struct_student_t   1
METROPOLIS_STEP_OF_STRUCTURAL_EFFECTS:
ROUNDS_WITHIN_CYCLE TUNING_SKIP
1   10
NUMBER_OF_STRUCTURAL_EFFECTS
1
STRUCTURAL_EFFECTS: LINE_FROM_EFFECTS_SECTION
SAVE_SAMPLES?
2   y
RESIDUAL_PRIOR_(CO)VARIANCES
150
RANDOM_GROUP
1
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_animal   1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
animodelh.ped   300   0
(CO)VARIANCES
100
```

A.6. Multibreed animal model.

```
# Example 6 - Multibreed animal model
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
4000   2000   20
SEED
123
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
mbreed1.dat   4000
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
3
OBSERVATION(S)
4
WEIGHT(S)

EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT [EFFECT NESTED]
1   4000   cross    n
2   1   cov          n
3   1   cov          n
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous   1
RESIDUAL_PRIOR_(CO)VARIANCES
100
RANDOM_GROUP
1
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_an_mb   3
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
mbreed1.ped   4000   0   2
MULTIPLE_SIRES: MAXN_FULL  [FILE: NAME DIMENSION]
[ALPHA_PRIORS]
0
METROPOLIS_STEP_OF_MULTIBREED_(CO)VARIANCES:
ROUNDS_WITHIN_CYCLE TUNING_SKIP
5   10
(CO)VARIANCES
100   20
 20   50
```

A.7. Reaction norms model.

```
# Example 7 - Reaction norms model
MCMC_CHAIN: TOTAL_CYCLES BURN_IN THINNING_INTERVAL
20000   5000   30
SEED
123
RESTART: Y/N? [CYCLE_TO_RESTART]
n
DATAFILE NAME N_RECORDS
rnorm_ex1.dat   1000
NUMBER_OF_TRAITS
1
NUMBER_OF_EFFECTS
4
OBSERVATION(S)
4
WEIGHT(S)
EFFECTS: POSITIONS_IN_DATAFILE NUMBER_OF_LEVELS
TYPE_OF_EFFECT [EFFECT NESTED]
2   1   cov           n
3   50   unknowcov          n
2   1520   cov      n   1
3   1520   rnorm   n   1
RANDOM_RESIDUAL: TYPE PRIOR_DEGREES_OF_BELIEF
homogeneous   1
RESIDUAL_PRIOR_(CO)VARIANCES
100
RANDOM_GROUP
2
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
diagonal   1
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
(CO)VARIANCES
50
RANDOM_GROUP
3   4
RANDOM_TYPE PRIOR_DEGREES_OF_BELIEF
add_animal   2
PEDIGREEFILE: NAME N_ANIMAL N_GENETIC_GROUPS [N_BREEDS]
rnorm_ex1.ped   1520   0
(CO)VARIANCES
20   6
6    2
```

Embrapa
Pecuária Sul

CGPE 9143

Ministério da
Agricultura, Pecuária
e Abastecimento

BRASIL
UM PAÍS DE TODOS
GOVERNO FEDERAL