

**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA”
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CAMPÔS DE JABOTICABAL**

**MODELOS DE ANTEDEPENDÊNCIA PARA AVALIAÇÃO
GENÉTICA DE CARACTERÍSTICAS LONGITUDINAIS**

FRANCISCO RIBEIRO DE ARAUJO NETO
Zootecnista

JABOTICABAL – SÃO PAULO - BRASIL
2012

**UNIVERSIDADE ESTADUAL PAULISTA “JULIO DE MESQUITA”
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CAMPÔS DE JABOTICABAL**

**MODELOS DE ANTEDEPENDÊNCIA PARA AVALIAÇÃO
GENÉTICA DE CARACTERÍSTICAS LONGITUDINAIS**

FRANCISCO RIBEIRO DE ARAUJO NETO
Orientador: Prof. Dr. HENRIQUE NUNES DE OLIVEIRA

**Tese apresentada à Faculdade de Ciências
Agrárias e Veterinárias – Unesp, Câmpus de
Jaboticabal, como parte das exigências para a
obtenção do título de Doutor em Genética e
Melhoramento Animal.**

JABOTICABAL – SP
JULHO DE 2012

A663m Araujo Neto, Francisco Ribeiro
Modelos de antedependência para avaliação genética de características longitudinais / Francisco Ribeiro de Araujo Neto. -- Jaboticabal, 2012
vii, 47 f. ; 28 cm

Tese (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrárias e Veterinárias, 2012
Orientador: Henrique Nunes de Oliveira
Banca examinadora: Lucia Galvão de Albuquerque, Danísio Prado Munari, Newton Tamassia Pégolo, Lenira El Faro Zadra
Bibliografia

1. Antedependência. 2. Amostrador de Gibss. 3. Coeficientes autorregressivos. 4. Medidas repetidas. 5. Variâncias das inovações. I. Título. II. Jaboticabal-Faculdade de Ciências Agrárias e Veterinárias.

CDU 636.082:519.2

DADOS CURRICULARES DO AUTOR

FRANCISCO RIBEIRO DE ARAUJO NETO – nascido a 30 de julho de 1983, na cidade de Rio Verde - Goiás, é formado em Zootecnia pela FESURV - Universidade De Rio Verde – RS, em julho de 2005 e, Mestre em Genética e Melhoramento Animal pela Universidade Estadual Paulista, sob orientação do Prof^o. Henrique Nunes de Oliveira. Em agosto de 2008, iniciou o doutorado em Genética e Melhoramento Animal, na Faculdade de Ciências Agrárias e Veterinárias, UNESP, Campus de Jaboticabal, continuando sob a orientação do Prof. Dr. Henrique Nunes de Oliveira.

Dedico.....

À Deus
À meus pais.
À meus irmãos.

AGRADECIMENTOS

Agradeço primeiramente a Deus, uma vez que sem sua vontade nada é possível e por acreditar que nenhum homem seja capaz de superar obstáculos sem fé, seja qual for a sua crença.

À meus pais, Jerônimo Vieira de Araújo e Terezinha Ferreira dos Anjos, cujos sonhos e objetivos sempre estiveram ligados à mim e a meus irmãos, demonstrando alegria nos momentos felizes, esperança e apoio em momentos difíceis.

À meus irmãos, Stefane Ferreira de Araújo e Jerônimo Vieira de Araújo Filho, que apesar de nossas diferenças, sempre estiveram a meu lado seja com palavras de incentivo, ou com apenas um ombro amigo.

Ao professor Dr. Henrique Nunes de Oliveira, pelos ensinamentos e orientação.

À professora Dr^a. Lúcia Galvão de Albuquerque, à Dr^a. Lenira El Faro Zadra, ao Dr. Danisio Prado Munari e ao Dr. Newton Tamassia Pegolo por comporem a banca de defesa, contribuindo com sugestões que aprimoraram este trabalho.

Àos colegas de pós-graduação: Agmar, Tammya, Severino, Dimas, Carla, Davi, Joseane, Ana Flávia, Lina Maria, Henry, Raul, Breno, Arione, Juana, Denise, Rodrigo Junqueira, Leonardo, Naudim, Fabio Borba, Alexandre, Marina, Daniel Gordo, Daniel, Daniele Portela, Matilde, Diogo, Inae e Iara pelo companheirismo.

À amigos de longa data: Cassia, Juarez, Marcia, Ana Paula, Milene, Idaiana, Odilon e Cristiane, entre tantos, que sempre estiveram próximos a mim.

Ao programa de pós-graduação em genética e melhoramento animal da UNESP, pela possibilidade oferecida de desenvolver o mestrado.

A ANCP pela cessão dos dados, à Fundação de Amparo a Pesquisa do Estado de São Paulo - FAPESP e à Capes pela concessão de bolsas, que permitiram o desenvolvimento deste trabalho.

SUMÁRIO

1	INTRODUÇÃO.....	1
2	OBJETIVOS.....	2
2.1	OBJETIVO GERAL.....	2
2.2	OBJETIVOS ESPECIFICOS.....	2
3	REVISÃO DE LITERATURA.....	3
3.1	Metodologias de estudo de dados longitudinais.....	3
3.2	Modelos de antedependência.....	5
3.3	Inferência bayesiana.....	7
4	MATERIAIS E METODOS.....	10
4.1	O modelo Geral.....	10
4.2	Inferência bayesiana.....	12
4.3	Implementação computacional.....	16
4.4	Estudos de simulação.....	17
4.4.1	Simulação dos registros de dados.....	18
4.4.2	Análises de antedependência.....	20
5	RESULTADOS E DISCUSSÃO.....	21
6	CONCLUSÃO.....	29
	REFERÊNCIAS.....	30
	APENDICE.....	39

LISTA DE FIGURAS

FIGURA 1	Estimativas de variâncias de dados simulados (Estudo 1-AD) utilizando de diferentes modelos de antedependência.....	21
FIGURA 2	Estimativas de variâncias de dados simulados (Estudo 1-MRA) utilizando de diferentes modelos de antedependência.....	22
FIGURA 3	Estimativas de variâncias de dados simulados (Estudo 2) utilizando de diferentes modelos de antedependência.....	23
FIGURA 4	Representação gráfica das cadeias de monte Carlo, para os parâmetros autorregressivos, estimados para os efeitos genéticos e de ambiente permanente para as replicas simuladas.....	25

LISTA DE TABELAS

TABELA 1	Parâmetros empregados para a simulação de dados para dois estudos: a) estudo 1-efeito genético e residual utilizando de estrutura de antedependência (AD) e regressão aleatória (MRA) e; b) estudo 2-simulação para um modelo com efeito genético, de ambiente permanente e residual.....	18
TABELA 2	Número de amostras efetivas obtidas para cada parâmetro estimado para o modelo de antedependência no segundo estudo de simulação em todas as réplicas.....	24

MODELOS DE ANTEDEPENDÊNCIA PARA AVALIAÇÃO GENÉTICA DE CARACTERÍSTICAS LONGITUDINAIS

RESUMO: Em programas de melhoramento animal, diversas características utilizadas como critérios de seleção representam medidas repetidas do mesmo animal, sendo importante o estudo de metodologias que considere esta estrutura longitudinal. Assim, o objetivo deste trabalho foi implementar a metodologia de antedependência utilizando de inferência bayesiana, na estimação de componentes de variância para avaliação genética de características longitudinais. Foram realizados dois estudos de simulação a fim de verificar as propriedades da aplicação da metodologia. Para o primeiro estudo foram considerados apenas os efeitos fixos, o genético e resíduo, sendo empregado duas estruturas de (co)variância: uma de antedependência e outra de regressão aleatória. O segundo estudo apresenta adicionalmente o efeito de ambiente permanente, sendo utilizado para ambos os efeitos aleatórios. Os resultados demonstraram que a metodologia proposta apresentou capacidade de estimar valores próximos aos simulados, entretanto alguns parâmetros dos modelos apresentaram algumas dificuldades de convergência. Verificou-se por meio do estudo realizado, que a implementação da metodologia de antedependência utilizando de inferência bayesiana e amostrador de Gibbs em um modelo animal, possibilitou a obtenção de estimativas próximas aos valores simulados. Além disso, pode-se constatar algumas dificuldades inerentes a convergência dos parâmetros do modelo, sendo necessário a realização de novos estudos visando utilizar novas funções para a modelagem dos coeficientes autorregressivos e das variâncias das inovações, assim como outras estratégias de amostragem.

Palavras-chaves: antedependência, amostrador de Gibbs, coeficientes autorregressivos, medidas repetidas, variância das inovações,

ANTEDEPENDENCE MODELS FOR GENETIC EVALUATION OF LONGITUDINAL TRAITS

ABSTRACT: In animal breeding programs, several traits used as selection criteria represent repeated measurements of the same animal, thus the study of methodologies that consider this longitudinal structure. The objective of this study was to implement a antedependence model using Bayesian inference, in the estimation of variance components for genetic evaluation of longitudinal characteristics. Two studies were conducted simulation in order to verify the properties of application of the methodology. For the first study included only fixed effects, genetic and residual, being employed two structures (co)variance: one antedependence and other random regression. The second study further shows the effect of permanent environment, and is used for both random effects. The results demonstrated that the proposed methodology showed ability to estimate values close to those simulated, however, some parameters of the models presented convergence difficulties. It was found through the study, the implementation of the methodology antedependence using Bayesian inference and Gibbs sampling in an animal model, makes possible to get estimates close to the simulated values. Moreover, one can observe some convergence difficulties of the model parameters, it is necessary to conduct new studies to use new functions for modeling the autoregressive coefficients and variances of the innovations, as well as other sampling strategies.

Keywords: antedependence, Gibbs sampling, autorregressives coefficients, repeated measures, innovation variance

1 - INTRODUÇÃO

Diante da atual ordem econômica, fundamentada na competição e na busca pela eficiência produtiva, a adoção de novas tecnologias pelos produtores constitui-se uma necessidade. Estas mudanças no manejo dentro das propriedades são acompanhadas por alterações genéticas nos rebanhos, promovidas pela seleção de animais. Este processo encontra-se apoiado na escolha de critérios condizentes com o objetivo da atividade e na estimação acurada de componentes de (co)variância das características a serem selecionadas.

A estimação de componentes de (co)variância para quaisquer critérios adotados permite a construção e emprego de índices de seleção e das equações de modelos mistos, a obtenção de parâmetros como herdabilidade e correlação genética, permitindo o planejamento de programas de melhoramento e a interpretação de mecanismos genéticos envolvidos na expressão das características (HENDERSON, 1986). Para tanto, deve-se estabelecer um modelo que possibilite descrever de forma satisfatória os registros das características.

Empiricamente, pode-se constatar que o crescimento de um animal se constitui em uma série de medidas distribuídas ao longo do tempo, o que rotula estas informações como longitudinais, ou ainda como “características de dimensão infinita” (KIRKPATRICK et al., 1990). Afora o crescimento, diversas são as características econômicas que assim se caracterizam, tais como perímetro escrotal, intervalos de parto, produção de leite, dentre outras. Este tipo de atributo merece um tratamento estatístico distinto, em virtude de um padrão de covariâncias bem estruturado. Necessita-se, portanto, do desenvolvimento de novas metodologias que possibilitem a modelagem e a análise dos dados longitudinais de forma a maximizar a acurácia das avaliações, tornando o processo de seleção mais eficiente e, proporcionando um incremento dos ganhos genéticos advindos deste.

2 - OBJETIVOS

2.1 -OBJETIVO GERAL

- Implementação da metodologia de antedependência em modelos de análise genética de dados longitudinais.

2.2 - OBJETIVOS ESPECIFICOS

- Adaptação de distribuições posteriores condicionais para estimação de parâmetros de modelos de antedependência em estudos genético quantitativos;
- Desenvolvimento de sub-rotinas e de um programa computacional para a aplicação da metodologia de antedependência utilizando métodos de Monte Carlo e Cadeias de Markov (MCMC);
- Verificar a capacidade de modelagem de diferentes estruturas de variâncias e aspectos relacionados a cadeias de MCMC por meio de estudos de simulação.

3 - REVISÃO DE LITERATURA

3.1 – Metodologias de estudo de dados longitudinais

A princípio, a forma de análise mais simples de ser efetuada no estudo de dados longitudinais consiste na utilização de modelos de repetibilidade, pressupondo-se que os registros repetidos são expressões de uma mesma característica. Portanto considera-se que as correlações entre as medidas sejam iguais à unidade, e que a (co) variância entre todos os pares de registros são de mesma magnitude, situação esta que não corresponde à realidade biológica das características.

Uma alternativa consiste em considerar cada medida como uma característica diferente e realizar uma análise conjunta em um modelo multi-características, sendo esta uma abordagem freqüente em análises para diversas características de interesse (MUCARI & OLIVEIRA, 2003; SARMENTO et al., 2003; SIQUEIRA et al.; 2003; PEREIRA et al., 2006; FARIA et al., 2007). Comparada à análise anterior, esta possui a vantagem de oferecer diferentes estimativas de correlação entre as medidas, mas sem assumir qualquer estrutura de covariância. Entretanto, dependendo do número de medidas envolvidas na análise, é necessária a estimação de um grande número de parâmetros, aumentando o custo computacional.

Outra forma de estudo é através do emprego de funções não-lineares, seja para a estimação de curvas de crescimento (OLIVEIRA, 1995; MENDES, 2007), ou de lactação. Tem-se em uma fase inicial o ajuste de curvas individuais e, posteriormente, os parâmetros obtidos são decompostos em seus componentes genéticos e ambientais, sendo a seleção realizada sobre estes. No entanto, tal procedimento raramente tem sido utilizado em programas de melhoramento de gado de corte, em virtude da necessidade de um grande número de mensurações para minimizar os erros de ajuste dos parâmetros do modelo considerado.

Nos últimos anos, a metodologia de regressão aleatória (MRA) tem sido adotada como procedimento padrão no que se refere a dados longitudinais em melhoramento animal. Encontrando-se atualmente estabelecida a metodologia, tem sido empregada em diversas espécies domésticas, como bovinos de corte (ALBUQUERQUE & MEYER, 2001; DIAS et al., 2006); bovinos de leite (EL FARO & ALBUQUERQUE, 2003; ARAÚJO et al., 2006).

A teoria associada à regressão aleatória foi apresentada inicialmente por Henderson Junior (1982). O método se fundamenta na modelagem de uma curva de regressão fixa (tendência média da população) e, em equações de regressão para cada efeito aleatório do modelo. Assume-se uma estrutura de covariância entre os coeficientes de regressão, representada por uma função de covariâncias (FC), que adota polinômios ortogonais como covariáveis, sendo os mais utilizados os polinômios de Legendre (SCHAEFFER, 2004).

Dentre as deficiências dos MRA - ocasionadas, principalmente, pela utilização dos polinômios de Legendre – tem-se a baixa capacidade de modelagem de funções assimétricas, sensibilidade da curva de (co) variância em todos os parâmetros, e em casos de modelos mais complexos, utilização de polinômios de ordens altas, implicando em uma hiper-parametrização (MISZTAL, 2006). Inúmeras alternativas têm sido propostas para contornar estes problemas, tais com a utilização de “b-splines” (MEYER, 2005) e de outros tipos de funções, como as de Jacobi modificadas, trigonométricas e de “Ali e Shaeffer” (COBUCI et al., 2004; MACEDO, 2007).

Por fim, outra opção para a análise de dados longitudinais consiste na modelagem paramétrica das matrizes de (co)variância (GEARY, 1989; JENNRICH & SCHLUCHTER, 1986 e WOLFINGER, 1996), sendo que as funções utilizadas com esta finalidade devem apresentar como requisitos básicos: a) ser positiva semi-definida; b) uma vez que o processo biológico apresente uma trajetória suave, a função deve apresentar este mesmo caráter e; c) é desejável, mas não obrigatório, que apresente parâmetros de fácil interpretação biológica, de forma

que hipóteses propostas possam ser facilmente testadas (PLETCHER & GEYER, 1999).

Em estudos na área de melhoramento animal, verifica-se que a aplicação desta metodologia se dá, em sua maioria, na modelagem de resíduos correlacionados em estudos de curvas de crescimento não lineares (PALA et AL., 2005; MENDES, 2007) e, no caso de estimação de componentes de (co) variância, a maioria dos estudos limita sua utilização apenas ao efeito de ambiente permanente do animal (CARVALHEIRA et al., 1998; ALBUQUERQUE & MEYER, 2005). Constata-se, portanto, um pequeno número de trabalhos em que foi empregada a parametrização para a estimação de componentes genéticos (JAFFRÉZIC & PLETCHER, 2000; JAFFRÉZIC et al., 2004), sendo utilizados em muitos deles as funções de antedependência estruturada.

3.2 – Modelos de antedependência

Os modelos de antedependência (GABRIEL, 1962), baseiam-se na dependência de uma determinada observação com outra precedente, sendo que, dependendo do tipo de função utilizado para estabelecer esta relação, diferentes especificações podem ser obtidas com diferentes propriedades estatísticas. Desta forma, considerando uma especificação autorregressiva da teoria, temos as seguintes equações para medidas de um animal, no tempo j (y):

$$y_1 = x_1' \beta + \delta_1$$

$$y_j = x_1' \beta + \sum_{k=1}^{s^*} \varphi_{jk} (y_{j-k} - x_{j-k}' \beta) + \delta_j$$

onde $x_1' \beta$ representa efeitos fixos; k representa a ordem de antedependência empregada, ($k=1, \dots, s$); o termo δ_j apresenta-se como uma variável aleatória normal, denominada de inovação distribuída com média zero e variância tempo-dependente σ_j^2 , e os φ_{jk} são parâmetros autorregressivos.

Os modelos antedependentes, à semelhança dos estacionários autorregressivos, levam em conta correlação serial, mas os primeiros são mais gerais no sentido de que eles não estipulam que todas as variâncias sejam iguais e nem que todas as correlações entre pares de medidas (no mesmo indivíduo) equidistantes no tempo são as mesmas, além de permitir que os intervalos de tempo entre as observações sejam diferentes, tal como em modelos de correlação de Markov (CROWDER & HAND, 1990; ZIMMERMAN et al., 1998).

Mesmo que o número de parâmetros a serem estimados seja menor que em outras análises (como em modelos multi-características), os modelos de antedependência podem ainda apresentar um número considerável de parâmetros. Assim, visando aumentar a parcimônia dos modelos de antedependência, NÚÑEZ-ANTÓN e ZIMMERMAN (2000) propuseram modelar os parâmetros de antedependência utilizando funções contínuas, permitindo inclusive procedimentos com dados desigualmente espaçados, bem como dados desbalanceados.

No que concerne às variâncias das inovações, a utilização de uma função polinomial poderá, assim como nos modelos de regressão aleatória, apresentar um efeito de borda. No entanto, a relevância deste fenômeno apresenta-se menor nos modelos SAD, uma vez que as funções de covariância estimadas são resultado não somente destes polinômios, mas de uma combinação complexa de parâmetros de antedependência e variância das inovações (JAFFRÉZIC et al., 2004).

De acordo com MACCHIAVELLI & MOSER (1997) e POURAHMADI (1999), a relação entre os parâmetros de antedependência descritos e a matriz de (co)variância dos registros (G), pode ser deduzida pela definição de uma matriz diagonal D com as variâncias das inovações e, uma matriz triangular inferior C , que apresenta como elementos, 1 na diagonal principal e valores negativos dos parâmetros autorregressivos nas sub-diagonais, de acordo com a ordem do modelo utilizado, de forma que:

$$G^{-1} = CD^{-1}C^T$$

onde, a expressão é proporcional a decomposição de Cholesky modificada, da matriz de (co)variância.

No que se a aplicação na estimação de efeitos genéticos poucos estudos podem ser descritos. JAFFRÉZIC et al. (2003) realizaram estudo onde se buscou comparar o modelos de antedependência estruturada com outras metodologias de predição em dados longitudinais, dentre elas o modelo de regressão aleatória, em análises uni e bi-característica, sendo utilizados dois bancos de dados: a) informações de fertilidade e mortalidade em *Drosophila melanogaster* e; b) produção de leite, proteína e gordura em vacas de leite. JAFFRÉZIC et al. (2004) aplicaram os modelos de antedependência estruturada no estudo genético de curvas de crescimento de bovinos da raça Charolês, sendo os resultados comparados com os obtidos via modelos de regressão aleatória.

Em relação aos métodos de estimação, grande parte dos estudos utilizando esta função utilizam de abordagens frequentistas (máxima verossimilhança ou máxima verossimilhança restrita), entretanto trabalhos empregando inferência bayesiana e métodos de Monte Carlo tem sido realizado com modelos simples (CEPEDA-CUERVO & GAMERMAN, 2004; CEPEDA-CUERVO & NUNEZ-ANTON, 2007).

3.3 - Inferência bayesiana

A análise bayesiana consiste em uma metodologia de inferir, no qual os parâmetros são tidos como variáveis aleatórias detentoras de distribuições “*a priori*” que refletem o estágio de conhecimento sobre eles, antes mesmo de se obterem os dados (GIANOLA & FERNANDO, 1986; BEAUMONT & RANNALA, 2004). Neste contexto, objetiva-se descrever toda a informação sobre o valor de determinado parâmetro, utilizando-se da probabilidade como uma medida condicional da incerteza (BLASCO, 2001; BERNARDO, 2002).

Os métodos bayesianos encontram-se fundamentados no teorema desenvolvido pelo reverendo Thomas Bayes, em 1761, o qual se compõe em uma forma matemática de resumir toda a informação disponível sobre um determinado evento, com a finalidade de realizar a melhor decisão possível (BERNARDO & SMITH, 1993). Através de inúmeras considerações, a expressão base da inferência bayesiana pode ser representada pela seguinte proporcionalidade:

$$p(\theta|y) \propto p(y|\theta) p(\theta)$$

em que, $p(\theta)$ é a densidade de probabilidade “*a priori*” de θ , a qual reflete o grau de conhecimento acumulado sobre os possíveis valores de θ , antes da obtenção de informações contidas em y ; $p(y|\theta)$ corresponde a uma função de verossimilhança de y , que representa a contribuição dos dados ao conhecimento do parâmetro; e $p(\theta|y)$ é a densidade de probabilidade “*a posteriori*” de θ , sobre a qual são realizadas as inferências sobre determinado parâmetro de interesse, através da determinação de estimadores pontuais (média, mediana e moda) e por intervalares (intervalos de credibilidade bayesiano e de alta densidade).

Por intermédio da expressão base, pode-se verificar uma atualização das informações contidas na distribuição “*a priori*” de θ , pela função de verossimilhança. Deste modo, a distribuição “*a posteriori*” contempla o grau de conhecimento prévio sobre o parâmetro $p(\theta)$ e também as informações adicionais propiciadas pelo experimento $p(y|\theta)$ (RESENDE, 2000; BEAUMONT & RANNALA, 2004). Um ponto de relevância é a influência do volume de dados, sobre a ponderação da informação “*a priori*” e da verossimilhança na determinação da “*posteriori*”. Em geral, atribui-se uma maior importância a primeira quando a quantidade (n) de dados experimentais atuais é pequena. À medida que o tamanho da amostra aumenta, a $p(\theta|y)$ tende a se assemelhar a distribuição representada pela verossimilhança (GIANOLA & FERNANDO, 1986; CARNEIRO JUNIOR et al., 2005).

Uma das principais características da metodologia Bayesiana, que a diferencia da clássica, é a abordagem conferida no momento de se fazer inferência estatística sobre os parâmetros, a partir de uma amostra. Nos procedimentos freqüentistas assumem-se que os parâmetros do modelo probabilístico são valores fixos ou constantes, embora desconhecidos, podendo existir vários estimadores. Na técnica Bayesiana, todavia, os parâmetros são vistos como variáveis aleatórias, existindo a princípio, um único estimador, o qual conduz a estimativas que maximizam uma função densidade de probabilidade “*a posteriori*” (RESENDE et al, 2001; TORRES et al., 2007).

Outra diferença é com respeito a alguns conceitos rotineiros em estatística, que não apresentam interpretação bayesiana. Em um contexto bayesiano não existe “viés” uma vez que, repetições de um experimento não são consideradas. Com relação à categorização dos efeitos, a princípio, todos são considerados aleatórios, tornando inócua a designação de modelos mistos. No entanto, a distinção requerida na genética quantitativa, é providenciada pela atribuição de “*prioris*” distintas entre os efeitos ditos “fixos” e “aleatórios” (GIANOLA & FERNANDO, 1986; BLASCO, 2001).

No que tange ao emprego de tais métodos em melhoramento animal, estes surgiram inicialmente como uma solução para a avaliação genética em características de limiar (GIANOLA & FOLLEY, 1982, 1983); sendo a teoria posteriormente, expandida para características de manifestação contínua (GIANOLA & FERNANDO, 1986), e empregada em diversos estudos para estimação de componentes de variância e valores genéticos, nas mais diversas espécies domésticas (MAGNABOSCO et al., 2000; MAGNABOSCO et al., 2001; NOGUEIRA et al., 2003; BARBOSA, 2005; GARA et al., 2006; MASCIOLI et al., 2006; MELLO et al., 2006; WINTER et al., 2006; YAMAKI, 2006; TORAL et al., 2007).

4 – MATERIAIS E METODOS

4.1 - O modelo geral

As equações de um modelo linear misto base para a análise de dados longitudinais podem ser representadas por:

$$y = Xb + Za + Wc + e$$

onde y representa o vetor das observações; a e c são os vetores de soluções para os efeitos genético direto e ambiente permanente direto, respectivamente; b e e são nesta ordem, os vetores de efeitos fixos e dos resíduos; X , Z e W são matrizes de incidência que relacionam os vetores b , a e c , respectivamente, ao vetor de observações.

Para o estudo de antedependência foi empregada a caracterização autorregressiva, de forma que para um determinado nível i do g -ésimo efeito aleatório do modelo, para um animal, tem-se que:

$$g_i = \sum_{j=1}^{i-1} \phi_{ij} g_j + v_i$$

onde Φ_{ij} representam os parâmetros autorregressivos entre medidas no tempo i e j e, v_i representa o erro do ajuste, denominada inovação, as quais são assumidas mutuamente independentes. Em termos matriciais temos:

$$g = Tg + v$$

em que T representa uma matriz com sub-diagonais contendo os coeficiente autorregressivos, sendo que v é distribuído com uma matriz diagonal Δ . Assim a matriz de (co)variância de g pode ser parametrizada como:

$$G = (I - T)^{-1} \Delta (I - T)$$

$$G = (T_p)^{-1} \Delta T_p$$

onde T_p consiste em uma matriz com 1 na diagonal e valores negativos dos coeficientes autoregressivos nas sub-diagonais.

Baseado no modelo misto e na parametrização de antedependência adotados, as seguintes pressuposições foram consideradas:

$$E \begin{bmatrix} y \\ a \\ c \end{bmatrix} = \begin{bmatrix} Xb \\ 0 \\ 0 \end{bmatrix} \quad V \begin{bmatrix} a \\ c \\ e \end{bmatrix} = \begin{bmatrix} A \otimes (H^{-1} D (H^{-1})^T) & 0 & 0 \\ 0 & I_{N_a} \otimes (K^{-1} J (K^{-1})^T) & 0 \\ 0 & 0 & R \end{bmatrix}$$

em que: H e K são as matrizes que apresentam os coeficientes autorregressivos para os efeitos genético aditivo direto e de ambiente permanente do animal; D e J são as matrizes diagonais das variâncias das inovações para os efeitos genético aditivo direto e de ambiente permanente do animal; A é a matriz de parentesco entre os indivíduos; I é a matriz de identidade; N_a é o número de animais com registro; \otimes é o produto de kronecker entre matrizes; e R representa variâncias modeladas através de classes heterogêneas.

Tendo em vista as considerações acima, as equações dos modelos mistos para o calculo do melhores estimadores e preditores lineares não viesados (BLUE e BLUP) são abaixo representadas:

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z & X'R^{-1}W \\ Z'R^{-1}X & Z'R^{-1}Z + A^{-1} \otimes ((H_a)^T D_a^{-1} H_a) & Z'R^{-1}W \\ W'R^{-1}X & W'R^{-1}Z & W'R^{-1}W + I \otimes ((K)^T J^{-1} K) \end{pmatrix} \begin{pmatrix} \hat{b} \\ \hat{a} \\ c \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \\ W'R^{-1}y \end{pmatrix}$$

as quais podem ser representadas como:

$$C\rho = r$$

4.2 – Inferência bayesiana

O emprego de métodos bayesianos na estimação de parâmetros em modelos de antedependência foi originalmente proposto por DANIELS e POURAHMADI (2002), sendo posteriormente aplicados em modelos não-estruturados (CEPEDA & NUNEZ-ANTON, 2007) e estruturados (CEPEDA & GAMERMAN, 2004; DANIELS & ZHAO, 2003), não sendo relatada aplicação para efeitos aditivos.

Neste trabalho, em termos de parâmetros de antedependência, inicialmente as variâncias das inovações foram modeladas em classes, tornando possível a utilização apenas do algoritmo de amostragem de Gibbs. Foram consideradas distribuições gaussianas para os efeitos fixos, genéticos, de ambiente permanente e parâmetros autorregressivos; para as variâncias das inovações e residual foi utilizada uma distribuição gama invertida (GIANOLA & FERNANDO, 1986; DANIELS & ZHAO, 2003).

Primeiramente, a distribuição posterior condicional dos efeitos fixos e aleatórios (ρ), dados os demais parâmetros (ELSE), pode ser representada através do núcleo de uma distribuição de densidade normal (GIANOLA & FERNANDO, 1986):

$$\rho | ELSE \sim N(C^{-1}Q^T R^{-1}y, C^{-1})$$

onde $Q=[x \ z \ w]$. No que se refere aos parâmetros do modelo de antedependência, a teoria de suas posteriores encontra-se baseada na da regressão linear com resíduos heterocedásticos. No que se refere aos coeficientes autorregressivos, deve-se atentar para o fato de que os parâmetros estimados são funções destes coeficientes, de forma que:

$$g_i = \sum_{j=1}^{i-1} (w_{ij}^T \gamma) g_j + v_i$$

em que w_{ij} representa um vetor de covariáveis contínuas e, γ representa um vetor de coeficientes de regressão (ZIMMERMAN & NÚÑEZ-ANTÓN, 2010). Considerando todas as observações do animal, temos que:

$$g = V\gamma + v$$

$$g = [V_1, V_2 \dots V_r]\gamma + v$$

$$V_i = \begin{bmatrix} w_{ij}^T g_j \\ \vdots \\ w_{ij}^T g_j \end{bmatrix}$$

Em geral, nas derivações apresentadas na literatura são considerados efeitos independentes, entretanto em termos de análises genéticas em que matrizes diferentes de identidades são utilizadas (matriz de parentesco aditiva ou de dominância), torna-se necessário generalizar a formulação, a fim de obter formas quadráticas apropriadas para derivação das posteriores para os parâmetros do modelo. Assim, considerando que g seja distribuído normalmente com média 0 e matriz de variância G , a seguinte forma quadrática é estabelecida em sua priori:

$$g^T(C^{-1} \otimes G^{-1})g$$

$$g^T(C^{-1} \otimes (T_p^T D^{-1} T_p))g$$

Neste ponto, para resolver o produto kronecker entre C e a matriz de variância decomposta, tem-se:

$$C^{-1} \otimes (T_p^T D^{-1} T_p) = \begin{bmatrix} c_{11}(T_p^T D^{-1} T_p) & \dots & c_{1n}(T_p^T D^{-1} T_p) \\ \vdots & \ddots & \vdots \\ c_{n1}(T_p^T D^{-1} T_p) & \dots & c_{nn}(T_p^T D^{-1} T_p) \end{bmatrix}$$

O resultado acima pode ser obtido através da multiplicação de matrizes blocos, de forma que:

$$C^{-1} \otimes (T_p^T D^{-1} T_p) = \begin{bmatrix} T_p^T & & \\ & \ddots & \\ & & T_p^T \end{bmatrix} \begin{bmatrix} c_{11} D^{-1} & \dots & c_{1n} D^{-1} \\ \vdots & \ddots & \vdots \\ c_{n1} D^{-1} & \dots & c_{nn} D^{-1} \end{bmatrix} \begin{bmatrix} T_p & & \\ & \ddots & \\ & & T_p \end{bmatrix}$$

$$C^{-1} \otimes (T_p^T D^{-1} T_p) = (I \otimes T_p^T)(C^{-1} \otimes D^{-1})(I \otimes T_p)$$

Sendo que a equação acima satisfaz as propriedades do produtor kronecker (HARVILLE, 1997), pode-se finalizar o desenvolvimento da forma quadrática:

$$g^T(I \otimes T_p^T)(C^{-1} \otimes D^{-1})(I \otimes T_p)g$$

$$(g - V\gamma)^T(C^{-1} \otimes D^{-1})(g - V\gamma)$$

em que C representa uma matriz que estrutura a relação entre os animais, como por exemplo uma identidade para efeitos ambientais ou uma matriz de parentesco

aditivo para efeitos genéticos. Assim, considerando a transformação da forma quadrática acima, e a teoria de estimação bayesiana em modelos de regressão (TIAO & ZELLNER, 1964; BOX & TIAO, 1973) utilizada para o desenvolvimento de modelos de antedependência em modelos fixos (CEPEDA & GAMERMAN, 2004), obtemos a seguintes posteriores condicionais:

A) Coeficientes autorregressivos

$$\pi(H(\theta)|ELSE) \sim NMV \left[\left(V_a^t (A^{-1} \otimes D^{-1}) V_a + \Omega_{\tau(\theta)}^{-1} \right)^{-1} \left(V_a^t (A^{-1} \otimes D^{-1}) a + \Omega_{\tau(\theta)}^{-1} m_{\tau(\theta)} \right), \left(V_a^t (A^{-1} \otimes D^{-1}) V_a + \Omega_{\tau(\theta)}^{-1} \right)^{-1} \right]$$

$$\pi(K(\theta)|ELSE) \sim NMV \left[\left(V_c^t J^{-1} V_c + \Omega_{K(\theta)}^{-1} \right)^{-1} \left(V_c^t J^{-1} c + \Omega_{K(\theta)}^{-1} m_{K(\theta)} \right), \left(V_c^t J^{-1} V_c + \Omega_{K(\theta)}^{-1} \right)^{-1} \right]$$

B) Variâncias das inovações

$$\pi(D_i|ELSE) \sim GI \left[\frac{n_i}{2}, \frac{n_i (a_i - V_a T(\theta))^t (A^{-1}) (a_i - V_a T(\theta))}{2n_i} \right]$$

$$\pi(J_i|ELSE) \sim GI \left[\frac{r_i}{2}, \frac{r_i (c_i - V_c K(\theta))^t (c_i - V_c K(\theta))}{2r_i} \right]$$

C) Variância Residual

$$\pi(\tau|ELSE) \sim GI \left[\frac{r}{2}, \frac{r (Y - Xb - Za - Wc)^t (Y - Xb - Za - Wc)}{2} \right]$$

em que n_i e r_i representam o número de animais na matriz de parentesco e com registros e; r o número total de registros, considerando variância residual homogênea; $\Omega_{\tau(\theta)}^{-1}$ e $\Omega_{K(\theta)}^{-1}$ são nesta ordem as variâncias “a priori” dos parâmetros autorregressivos para o efeito genético e de ambiente permanente; $m_{\tau(\theta)}$ e $m_{K(\theta)}$

são as medias “*a priori*” dos parâmetros autorregressivos para o efeito genético e de ambiente permanente, respectivamente.

4.3 – Implementação computacional

Para a aplicação da metodologia de antedependência descrita acima, um programa em linguagem Fortran 95 foi desenvolvido, utilizando-se de matrizes esparsas para maior eficiência computacional. Como anteriormente definido, foram empregados os método de Amostragem de Gibbs para todos os parâmetros a serem estimados. O amostrador de Gibbs se apresenta como uma técnica para a obtenção indireta de variáveis aleatórias de uma distribuição (marginal), sem a necessidade de se calcular sua densidade de probabilidade (CASELA & GEORGE, 1992), sendo um caso particular do algoritmo de Metropolis-Hastings.

Diferentes variantes deste método foram empregados, sendo que para os efeitos fixos foram utilizados um Gibbs-escalar enquanto para os efeitos aleatórios, as amostragens foram realizadas em blocos individuais por animal/efeito. A diferença de métodos de amostragem entre os efeitos deve-se ao fato das soluções para as diversas medidas (tanto para o efeito genético como para de ambiente permanente) apresentarem forte correlação entre si o que dificulta a convergência no caso de se utilizar a amostragem escalar.

Foram testadas duas modelagens para os parâmetros que definem os coeficientes autorregressivos (CAU). A primeira (M01) consiste em modelar estes parâmetros em função do intervalo de tempo entre duas mensurações, como em CEPEDA & GAMERMAN (2004). A segunda (M02) consistiu em modelar o CAU como um componente para cada par de medidas, sendo obtidas equações para cada ordem considerada (assumindo uma estrutura de regressão múltipla), como nos estudos de JAFFREZIC et al. (2004).

Ao se considerar M01, pode-se constatar que duas medidas são explicadas por uma medida anterior na mesma magnitude, desde que equidistantes; fato este não verossímil ao se considerar os diversos estágios do crescimento de um

animal. Por outro lado, em M02 tem-se a estimação de um maior número de parâmetros, sendo que nas análises de simulação (não demonstrado) apresentaram-se mais consistentes no que se refere à convergência das cadeias de Gibbs e estimativas.

A versão inicial do programa (APÊNDICE 1) apresenta-se formulada apenas para os efeitos genéticos e de ambiente permanente direto, sendo que todas amostras são armazenadas em seu arquivo de saída. As análises de convergência e correlação entre amostras devem ser realizadas em outros programas, tais com o Gibanal ou pacotes como BOA e CODA, da plataforma estatística R. Outra restrição do programa é que devem ser utilizados bancos de dados em que todos os animais possuem todas as medidas, uma vez que ainda não foram implementadas sub-rotinas para dados perdidos.

4.4 – Estudos de simulação

Foram desenvolvidos dois estudos de simulação a fim de se validar a metodologia e o programa elaborados, além de se analisar a metodologia em cenários específicos. O primeiro estudo de simulação tem por objetivo verificar a capacidade da metodologia em estimar as variâncias, considerando duas estruturas distintas: a) baseada em um modelo de AD de segunda ordem e; b) uma estrutura baseada em uma matriz de coeficientes aleatórios. Para o segundo estudo, englobava todos os efeitos do modelo longitudinal básico, simulados utilizando de modelo AD, baseando-se na estrutura de um pedigree real.

Com relação a estrutura do pedigree utilizado nas simulações, para o primeiro estudo de simulação, foi gerado um arquivo com 5000 animais, distribuídos em 10 gerações discretas, sendo que em cada geração 20 machos e 80 fêmeas, eram selecionados e acasalados de forma aleatória. Para o segundo estudo foi utilizado um pedigree real, contendo 8500 animais.

4.4.1 – Simulação dos registros de dados

Os modelos utilizados para realizar as simulações, estão representados abaixo:

$$\text{Estudo 1: } y_{ji} = m_i + fix + a_{ji} + e_{ji}$$

$$\text{Estudo 2: } y_{ji} = m_i + fix + a_{ji} + p_{ij} + e_{ji}$$

onde, y_{ij} é a medida do j -ésimo animal no i -ésimo tempo; m_i , a_{ij} e p_{ij} são nesta ordem a média do i -ésimo tempo, o valor genético e de ambiente permanente para a referida medida e; fix e e_{ji} são efeitos fixos (com 2 níveis) e resíduo. Todos os efeitos aleatórios, foram considerados distribuídos com média 0, e a matriz de variância determinada de acordo com a simulação, e com resíduos homogêneos. Detalhes das estruturas utilizadas para a simulação das variâncias dos controles estão representados na tabela abaixo.

TABELA 1 - Parâmetros empregados para a simulação de dados para dois estudos: a) estudo 1-efeito genético e residual utilizando de estrutura de antedependência (AD) e regressão aleatória (MRA) e; b) estudo 2-simulação para um modelo com efeito genético, de ambiente permanente e residual

Estudo 1	
Estrutura A (AD)	Estrutura b (MRA)
Autorregressivo	
• ordem 1 (Quadrático):1.30; -0.11; 0.01	1.715020 -0.11442 -0.28401
• ordem 2 (Constante):0.2	-0.11442 0.21838 0.02552
Variâncias das inovações:	-0.28401 0.02552 0.12156
• 30 (primeira inovação)	
• 10 (demais inovações)	
Variância residual=100	Variância residual=1
Estudo 2	
Genético	Ambiente permanente
Autorregressivos: 0.95	
Autorregressivos: 1.1	
Variâncias das inovações:	
Variâncias das inovações:	
• 50 (primeira inovação)	• 70 (primeira inovação)
• 20 (demais inovações)	• 10 (demais inovações)
Variância residual=100	

Em ambos os estudos de simulação, para os animais que compunham a população-base (animais com pais e mães desconhecidos), os valores genéticos foram obtidos pela amostragem de uma distribuição normal multivariada, com media 0 e matriz de (co)variância própria para cada estudo (descritas anteriormente). Quando os pais eram conhecidos, os valores genéticos eram obtidos mediante:

$$a_i = 0.5a_{si} + 0.5a_{di} + \Psi$$

onde a_i , a_{si} e a_{di} são, respectivamente, vetores de valores genéticos do i -ésimo animal, de seu pai e mãe para todas as medidas. O termo Ψ representa a segregação mendeliana, gerado com media 0 e variância:

$$\Sigma_{\Psi} = (0.5 - 0.25(F_{si} + F_{di})) * \Sigma_a$$

de forma que F representa o coeficiente de endogamia dos pais do animal i ; Σ_{Ψ} e Σ_a representam nesta ordem, as matrizes de (co)variância da segregação mendeliana e dos valores genéticos.

Quando o animal apresentava pai ou mãe desconhecido, procedeu-se a obtenção do valor genético do parental desconhecido como para a população-base e posteriormente, seguia-se o procedimento apresentado anteriormente. Para os efeitos residuais e de ambiente permanente, assim como para o efeito genético, de distribuições normais, sendo todo o procedimento desenvolvido utilizando-se da plataforma estatística R.

4.4.2 – Análises de antedependência

Para a realização das análises de antedependência, foram considerados os efeitos dos modelos descritos na simulação, sendo testadas diversas combinações entre ordem de antedependência e polinômios. Assim em termos matriciais os modelos de análise testados foram:

$$\text{Estudo 1} \quad y = Xb + Za + e$$

$$\text{Estudo 2} \quad y = Xb + Za + Wc + e$$

em que y , b , a , c e e representam nesta ordem, os vetores de observações, efeitos fixos, valores genéticos, valores de ambiente permanente e resíduo; X , Z e W são as matrizes de incidência que relacionam os efeitos fixos, genéticos e de ambiente permanente ao vetor de observações, respectivamente. Foram consideradas distribuições gaussianas para os efeitos fixos, genéticos, de ambiente permanente e parâmetros autorregressivos; para as variâncias das inovações e residual foi utilizada uma distribuição gama invertida.

O rotulo das análises realizadas seguem o seguinte padrão: ADX-SEQ, onde X refere-se a ordem de antedependência utilizada e, SEQ corresponde a uma seqüência de letras que descrevem o polinômio empregado para modelar os coeficientes autorregressivos em cada ordem: C-constante; L-linear e Q-quadrático.

Com referência ao primeiro estudo de simulação, em cada análise foram geradas cadeias com 120000 amostras, enquanto para a simulação com estrutura longitudinal, foi utilizada uma cadeia com 500000 amostras. A definição de burn-in, amostras efetivas e estatísticas descritivas das distribuições posteriores foram realizadas utilizando-se do pacote Boa, disponível no software R.

5 – RESULTADOS E DISCUSSÃO

Primeiramente, considerando os resultados do primeiro estudo de simulação (Figuras 1 e 2), pode-se constatar, de uma forma geral, a capacidade da metodologia proposta em estimar os componentes de variância para o efeito genético aditivo, sendo esta capacidade determinada de acordo com a estrutura de variância utilizada na simulação. Para a primeira estrutura simulada (Figura 1) mesmo em modelos simples (AD1C) os valores estimados foram similares aos verdadeiros, sendo que para segunda estrutura (Figura 2) foi necessário a utilização de modelos mais complexos para obter melhores aproximações. Para esta última estrutura, verificamos que o modelo mais simples (AD1C) conseguiu estabelecer a tendência da curva de variâncias, mas estimando de forma errônea os pontos de maior variância.

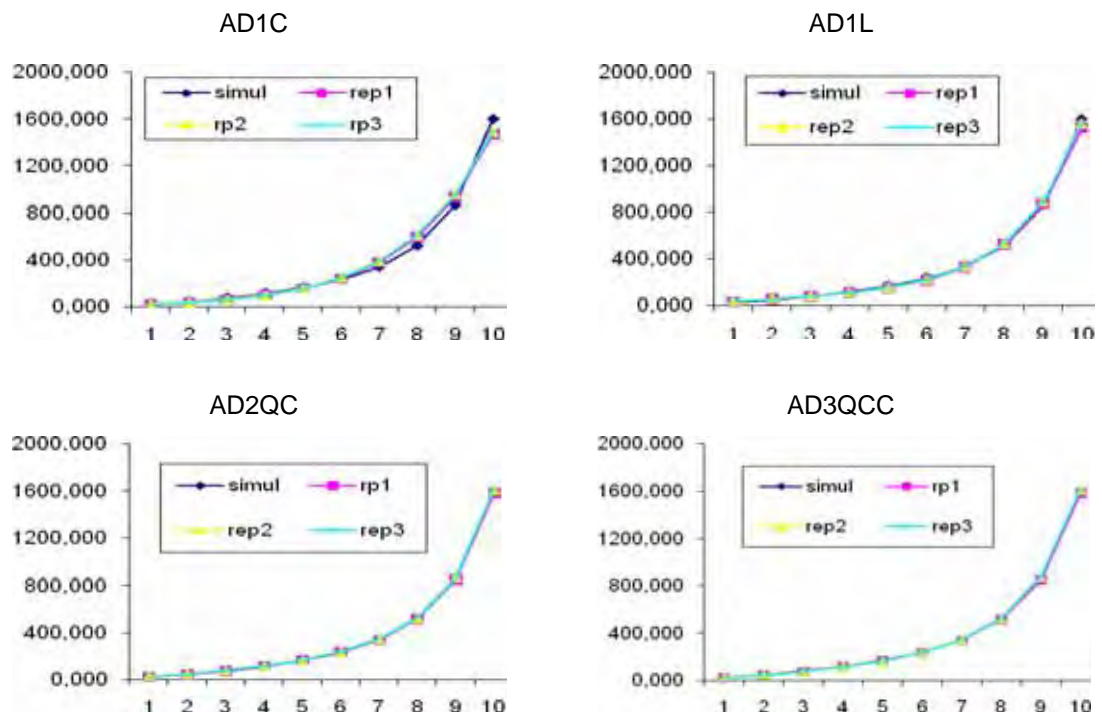


Figura 1 - Estimativas de variâncias de dados simulados (Estudo 1-AD) utilizando de diferentes modelos de antedependência.

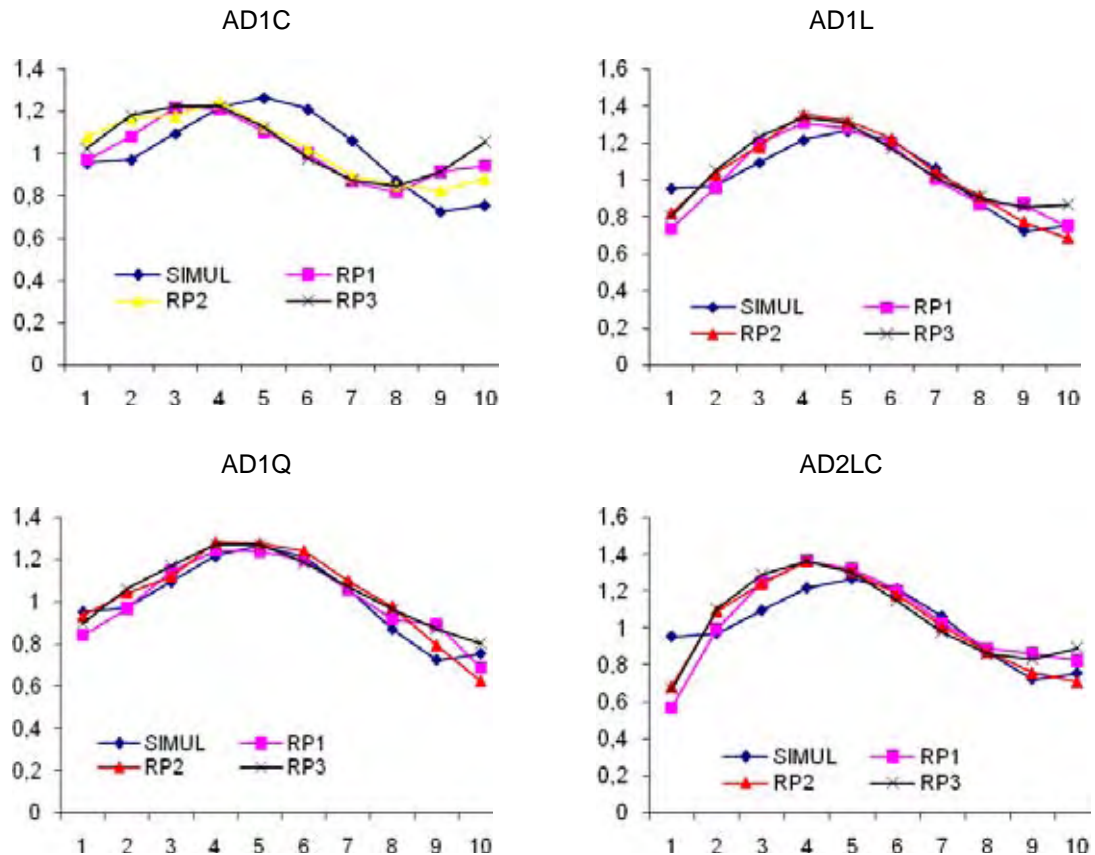


Figura 2 - Estimativas de variâncias de dados simulados (Estudo 1-MRA) utilizando de diferentes modelos de antedependência.

Além do fato da primeira estrutura ser simulada baseada em um modelo de antedependência, possivelmente outro fator que corroborou para facilitar a estimação foi o caráter crescente e regular das estimativas. Este tipo de comportamento pode ser verificado em estudos cujas características apresentam propriedade acumulativa, tal como pesos em diversas idades. Entretanto, para características como controles leiteiros ou ganhos de peso sejam necessários modelos mais complexos, uma vez que apresentam padrão de variância irregular (ARAUJO NETO et al., 2012; ASPILCUETA et al., 2012).

Ao se observar o estudo envolvendo o segundo padrão de variância, verificamos um aspecto particular dos modelos de antedependência (não verificado em modelos de regressão aleatória): os dois últimos modelos utilizados apresentam número igual de parâmetros a serem estimados, entretanto a

capacidade de estimação é diferente, em virtude das relações entre os parâmetros serem distintas. Assim verifica-se uma dificuldade inicial em modelos de AD no que se refere a seleção de modelos. A princípio, a função de verossimilhança inclusa na maioria dos critérios de seleção existentes, tais como BIC e AIC, deve ser capaz de representar estas diferentes relações entre parâmetros, dependendo do tamanho da amostra e número de efeitos modelados.

No que se refere a um modelo longitudinal -considerado no segundo estudo de simulação (Figura 3) – assim como no primeiro estudo, a estimação de valores foram próximos aos simulados em ambos os efeitos. Entretanto é possível verificar, que menor ajuste ocorre nas idades finais, em virtude da maior dificuldade de estimação das últimas variâncias das inovações; fenômeno este que pode ainda ser agravado com a utilização de um maior número de medidas. Cogita-se também que nesta situação, até mesmo encontrar funções para ajuste eficiente dos coeficientes autorregressivos pode se tornar difícil.

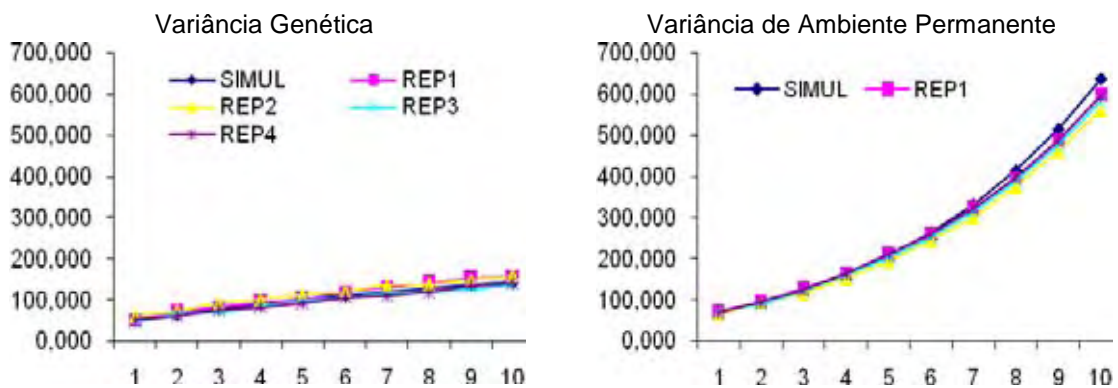


Figura 3 - Estimativas de variâncias de dados simulados (Estudo 2) utilizando de diferentes modelos de antedependência.

Esta dificuldade de estimar as variâncias das inovações reflete-se no número de amostras efetivas (Tabela 2), onde podemos notar uma tendência decrescente destas amostras em medidas finais, em virtude da auto-correlação entre as amostras.

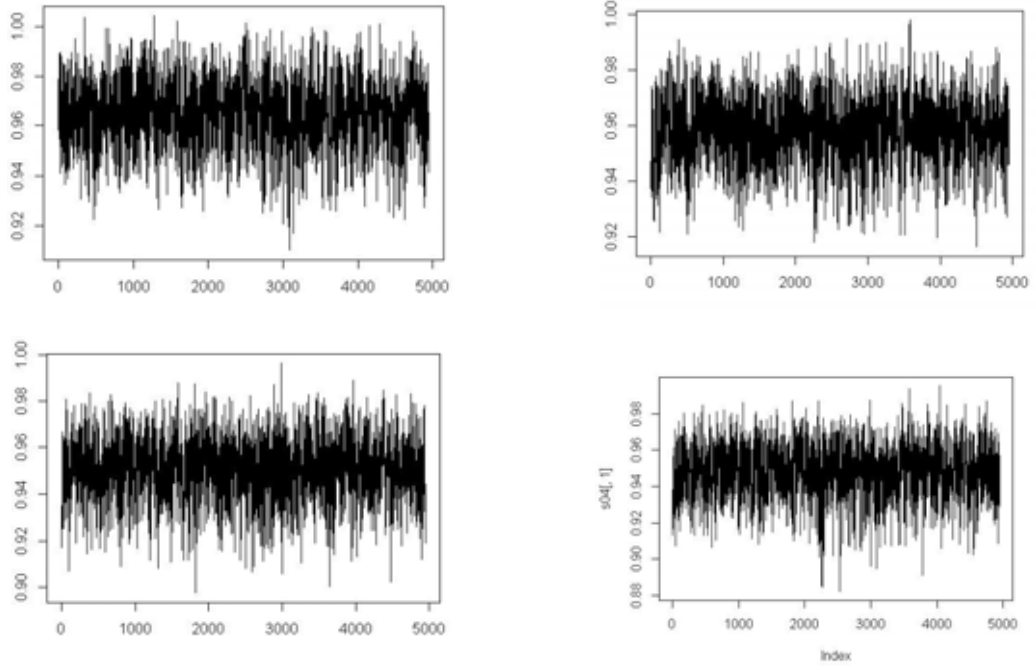
TABELA 2 – Número de amostras efetivas obtidas para cada parâmetro estimado para o modelo de antedependência no segundo estudo de simulação em todas as réplicas

Parâmetros*	rp1	Rp2	rp3	Rp4
AU_G	920,55	1059,25	1137,96	1059,25
VI_1_G	1706,08	1312,29	283,84	1312,29
VI_2_G	455,17	736,03	412,76	736,03
VI_3_G	821,77	1032,56	659,17	1032,56
VI_4_G	793,40	798,32	640,74	798,32
VI_5_G	817,56	1122,09	994,16	1122,09
VI_6_G	420,11	498,60	872,55	498,60
VI_7_G	872,16	526,39	903,43	526,39
VI_8_G	427,93	623,54	536,22	623,54
VI_9_G	748,99	421,77	263,83	421,77
VI_10_G	623,55	292,39	650,25	292,39
AU_A	807,06	694,70	82,47	694,70
VI_1_A	1619,60	968,81	322,00	968,81
VI_2_A	137,63	865,02	77,66	865,02
VI_3_A	667,86	694,87	831,76	694,87
VI_4_A	683,12	1046,30	663,67	1046,30
VI_5_A	790,36	783,15	494,52	783,15
VI_6_A	162,67	223,53	497,33	223,53
VI_7_A	564,24	335,22	660,83	335,22
VI_8_A	298,20	347,99	584,12	347,99
VI_9_A	371,85	225,99	48,05	225,99
VI_10_A	557,15	109,77	347,74	109,77
VRES	1973,29	1082,75	743,11	1082,75

*AU_X=parâmetro autorregressivo para o efeito X; VI_C_X=variância da inovação para a classe de idade C e para o efeito X; VRES=variância residual; G=efeito genético; A=efeito de ambiente permanente

Com relação à convergência do parâmetro autorregressivo (Figura 4), este apresentou algumas anomalias, principalmente para o efeito de ambiente permanente. Em geral, parâmetros associados a efeitos diagonais apresentam mais regulares que à outros associados com efeitos genéticos; entretanto pode-se considerar também que os valores iniciais fornecidos para os parâmetros do ambiente foram piores que aos fornecidos para o genético, o que causou maiores dificuldades de convergência.

Efeito genético aditivo



Efeito de ambiente permanente

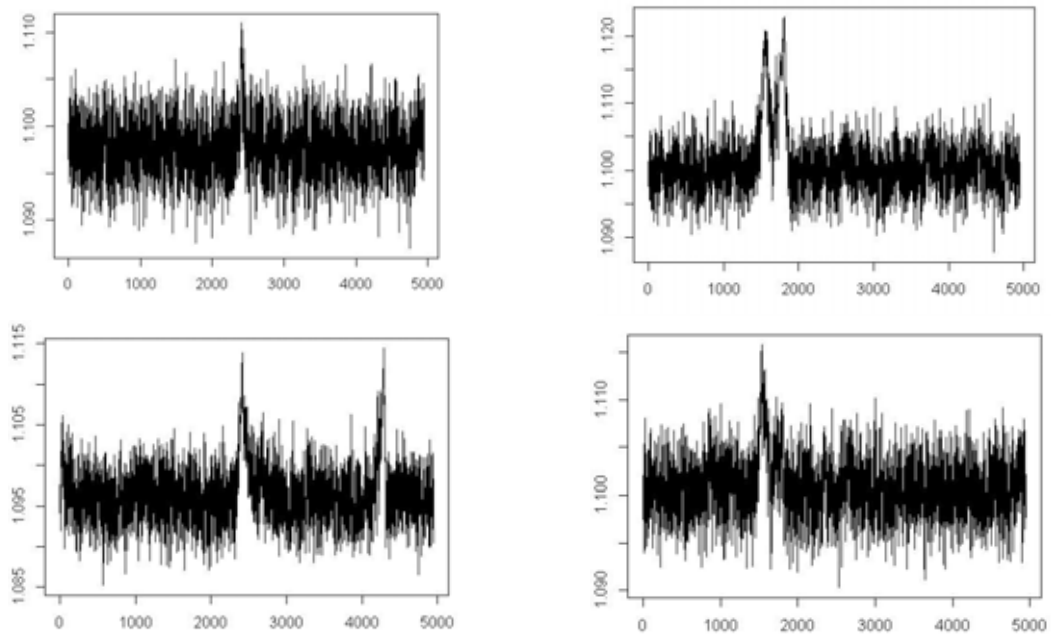


Figura 4 – Representação gráfica das cadeias de monte Carlo, para os parâmetros autorregressivos, estimados para os efeitos genéticos e de ambiente permanente para as replicas simuladas.

Um ponto inevitável quando do emprego de uma nova metodologia para estudo de dados longitudinais é a comparação com a metodologia de regressão aleatória (MRA). Podemos verificar primeiramente que diferente da MRA, não existe transformação do vetor de efeitos na metodologia de AD. Neste caso o número de equações para serem resolvidas durante o processo de estimação é maior que em MRA, porém com maior esparsidade.

Esta propriedade não somente possibilita menor consumo de memória RAM, mas também acelera o processo de resolução de equações, através da economia no número de operações. Assim, cada metodologia apresenta um aspecto particular favorável a implementação computacional. Possivelmente, ao se estabelecer uma relação entre número de medidas e esparsidade da matriz, poder-se-há verificar que, em termos computacionais, os modelos AD podem se apresentar mais interessantes que os de regressão aleatória quando da análise de poucas medidas (como controles mensais de leite), fato contrário sendo verificado em grandes dimensões.

A aplicação do método bayesiano se mostrou eficiente na estimação dos parâmetros do modelo AD, visando aplicação a efeitos aditivos. Entretanto, durante as análises verifica-se grande sensibilidade das análises aos valores dos parâmetros autorregressivos, que aqui foram considerados com variâncias fixas “a priori”. Tendo em vista a teoria da estimação bayesiana em modelos de regressão (base dos modelos antedependência), uma possibilidade que pode ser interessante é a aplicação de outros tipos de prioris. Em uma primeira opção, pode-se cogitar estimar a variância destes parâmetros em um modelo hierárquico (JOHNSON & JONES, 2010), o que pode aumentar o conteúdo informativo a cerca deste parâmetro e facilitar a convergência, mesmo com o aumento no número de parâmetros.

Outra opção interessante é aplicar lasso bayesiano (PARK & CASELLA, 2008), a fim de aumentar a eficiência na seleção de modelos (LYKOU & NTZOUFRAS, 2012). Como verificado ao se trabalhar com modelos AD devemos selecionar além da ordem, o grau de polinômio em cada ordem e a modelagem

das inovações, gerando um grande número de modelos. Ao serem adotadas *prioris* lasso pode-se estabelecer polinômios de alto grau para determinada ordem AD e, por conseguinte através das propriedades *shrinkage* os coeficientes de regressão pouco informativos tenderiam a zero.

Considerando a dificuldade de convergência de alguns parâmetros, principalmente as variâncias das inovações nas últimas idades, uma possibilidade é a utilização de esquemas alternativos de amostragem de Gibbs. A escolha da forma de amostragem interfere sobre a velocidade de convergência e a autocorrelação entre as amostras, e conseqüentemente, sobre a eficiência do processo de estimação. A utilização da amostragem em bloco (todos os valores genéticos simultâneos) acelera a convergência, e possibilita uma menor dependência entre as amostras, sendo que a amostragem escalar (cada valor genético amostrado individualmente) apresenta maior velocidade de execução e baixo consumo de memória. O que tem sido utilizado é a utilização de blocos intermediários, sendo que o esquema em blocos animal/efeito (utilizado na elaboração do programa) é adotado de uma forma geral como procedimento de amostragem quando analisadas várias características por animal (VAN TASSEL & VAN VLECK, 1996).

Tendo em vista a estrutura de parentesco, uma primeira opção pode ser a utilização de blocos de meio-irmãos, uma vez que na maioria das espécies esta seria a relação mais abundante dentro das estruturas populacionais, ou utilizando-se de linhas paternas. Outra possibilidade é a utilização de algoritmos híbridos, onde poderíamos utilizar amostragem de blocos maiores em intervalos pré-determinados (WALDMANN et.al, 2008) ou aplicar Metropolis-lavigne ou normais aproximadas para a amostragem dos efeitos em conjunto (WAAGEPETERSEN et al., 2008). A utilização desta última abordagem torna-se atrativa devido ao fato de não ser necessária a inversão da matriz de coeficientes (procedimento este com alta exigência computacional). Entretanto, sendo variantes do algoritmo de Metrópolis, baseiam-se na definição de uma distribuição aproximada, onde as amostras podem ser rejeitadas, o que implica na necessidade de cadeias maiores.

Com relação aos algoritmos híbridos, estes possuem a vantagem de não utilizar distribuições aproximadas, demonstrando nos estudos redução considerável da dependência das amostras: queda de 0,80 para 0,30 nas correlações seriais de primeira ordem (WALDMANN et.al, 2008). Mas, dependendo da definição dos blocos, dificuldades podem ser encontradas para sua implementação, tais como dificuldade de desenvolver sub-rotinas eficientes e aumento dos requerimentos computacionais.

6 - CONCLUSÃO

Verificou-se por meio do estudo realizado, que a implementação da metodologia de antedependência utilizando de inferência bayesiana e amostrador de Gibbs em um modelo animal, possibilitou a obtenção de estimativas próximas aos valores simulados. Além disso, pode-se constatar algumas dificuldades inerentes a convergência dos parâmetros do modelo, sendo necessário a realização de novos estudos visando utilizar novas funções para a modelagem dos coeficientes autorregressivos e das variâncias das inovações, assim como outras estratégias de amostragem.

REFERÊNCIAS

ALBUQUERQUE, L.G.; MEYER, K. Estimates of covariance functions for growth of Nelore cattle applying a parametric correlation structure to model within-animal correlations. **Livestock Production Science**, v.93, p. 213-222, 2005.

ALBUQUERQUE, L.G.; MEYER, K. Estimates of covariance functions for growth from birth to 630 days of age in Nelore cattle. **Journal Animal Science**, v.79, p. 2776-2789. 2001.

ARAÚJO, C.V.; TORRES, R.A.; COSTA, C.N.; TORRES FILHO, R.A.; ARAÚJO, S.I.; LOPES, P.S.; REGAZZI, A.J.; PEREIRA, C.S.; SARMENTO, J.L.R. Uso de modelos de regressão aleatória para descrever a variação genética da produção de leite na raça Holandesa. **Revista brasileira de Zootecnia**, v.35, n.3, p. 975-981, 2006 (supl).

BARBOSA, V. **Inferência bayesiana no estudo genético quantitativo de características de carcaça, utilizando a técnica de ultra-sonografia e suas relações com crescimento, em novilhos da raça Nelore**. 2005. 81 f. Dissertação (Mestre em Zootecnia) – Universidade Federal de Goiás, Goiânia, 2005.

BEAUMONT, M.A.; RANNALA, B. The bayesian revolution in genetics. **Nature Reviews/Genetics**, v.5, p.251-261.2004.

BERNARDO, J.M. Una introducció a l'estadística bayesiana. **Butlletí de La Societat Catalan de Matemàtiques**, v.17, n.1, 2002. p. 7-64.

BLASCO, A. The Bayesian controversy in animal breeding. **Journal of Animal Science**, v.79, p.2023-2046. 2001.

BOX, G.E.P.; TIAO, G.C. **Bayesian inference in statistical analysis**. New York: John Wiley and Sons, 1973. 588 p.

CARNEIRO JÚNIOR, J.M.; ASSIS, G.M.L.; EUCLYDES, R.F.; LOPES, P.S. Influência da informação *a priori* na avaliação genética animal utilizando dados simulados. **Revista Brasileira de Zootecnia**, v.34, n.6, p.1905-1913, 2005.

CARVALHEIRA, J.G.V.; BLAKE, R.W.; POLLAK, E.J.; QUAAS, R.L.; DURAN-CASTRO, C.V. Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucena cattle and in United States Holstein herds. **Journal of Dairy Science**, v.81, n.10, p.2738-2751. 1998.

CASELA, G.; GEORGE, E. I. **Explaining the Gibbs Sampler**. Technical Report No. Bu-1098-MA, Cornell University. 1992.

CEPEDA E.C; GAMERMAN, D. Bayesian modeling of joint regressions for the mean and covariance matrix. **Biometrical Journal**, v.46, p.430-440. 2004.

CEPEDA-CUERVO, E.; NUNEZ-ANTON, V. Bayesian joint modelling of the mean and covariance structures or normal longitudinal data. **SORT**, v.31,n2,p.181-200. 2007.

COBUCI,, J.A.; EUCLYDES, R.F.; COSTA, C.N.; LOPES, P.S.; TORRES, R.A; PEREIRA, C.S. Análises da persistência na lactação de vacas da raça Holandesa, usando produção no dia do controle e modelos de regressão aleatória. **Revista Brasileira de Zootecnia**, v.33, n.3., p.546-554, 2004.

CROWDER, M.J.; HAND, D.J. **Analysis of repeat measures**. London: Chapman & Hall, 1990. 257 p.

DANIELS, M.J.; ZHAO, Y.D. Modelling the random effects covariance matrix in longitudinal data. **Stat Med.**, v.22, n.10,p.1631-1647. 2003.

DIAS, L.T.; ALBUQUERQUE, L.G.; TONHATI, H.; TEIXEIRA, R.A. Estimação de parâmetros genéticos para peso do nascimento aos 550 dias de idade para animais da raça Tabapuã utilizando-se modelos de regressão aleatória. **Revista Brasileira de Zootecnia**, v.35, n.5, p.1915-125. 2006.

EL-FARO, L.;ALBUQUERQUE, L.G. Utilização de modelos de regressão aleatória para produção de leite no dia do controle, com diferentes estruturas de variâncias residuais. **Revista Brasileira de Zootecnia**, v.32, n.5.; p. 1104-1113, 2003.

FARIA, C.U.; MAGNABOSCO, C.U.; REYES, A.; LÔBO, R.B.; BEZERRA, L.A.F.; SAINZ, R.D. Bayesian inference in a quantitative genetic study of growth traits in Nelore cattle (*Bos indicus*). **Genetics and Molecular Biology**, v.30, n.3, p.545-551. 2007.

GABRIEL, K.R. Ante-dependence analysis of an ordered set of variables. **Ann. Math. Stat.**, v.33, p. 201-212. 1962.

GARA, A.B.; REKIK,B.; BOUALLÈGUE, M.; Genetic parameters and evaluation of the Tunisian dairy cattle population for milk yield by Bayesian and BLUP analyses. **Livestock Science**, v.100, p.142-149. 2006.

GEARY, D.N. Modelling the covariance structure of repeat measurements. **Biometrics**, v.45, n.4, p.1183-1195. 1989.

GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v.63, 1986, p.217-244.

GIANOLA, D.; FOULLEY, J.L. New techniques of prediction of breeding value for discontinuous traits. In: ANNUAL NATIONAL BREEDERS ROUNDTABLE, 32, St Louis, 1983. **Proceedings...**St. Louis: National Breeders Roundtable, 1983.p.128-154.

HENDERSON JUNIOR, C.R. Analysis of covariance in the mixed model: higher-level, nonhomogeneous and random regression. **Biometrics**, v.38, p.633-640. 1982.

HENDERSON, C.R. Recent developments in variance and covariance estimation. **Journal Animal Science**, v.63, p.208-216. 1986.

JAFFRÉZIC, F.; PLETCHER, S.D. Statistical models or estimating the genetic basis of repeated measures and other function-valued traits. **Genetics**, v.156, p.913-922. 2000.

JAFFRÉZIC, F.; THOMPSON, R.; HILL, W.G. Structured antedependence models for genetic analysis of repeat measures on multiple quantitative traits. **Genet. Res.**, Cambridge, v.82, p. 55-65. 2003.

JAFFRÉZIC, F.; VENOT, E.; LALOË, D.; VINET, A.; RENAND, G. Use of structures antedependence models for the genetic analysis of gro curves. **Journal Animal Science**, v.82, p. 3465-3473. 2004.

LYKOU, A.; NTZOUFRAS, I. On Bayesian lasso variable selection and the specification of the shirinkage parameter. **Stat Comput.**2012 (publicado on-line)

MACCHIAVELLI, R.E.; MOSER, E.B. Analysis of repeated measurements with ante-dependence covariance models. **Biometrical Journal**, v.39, n.3, p.339-350. 1997.

MACEDO, O.J. **Modelos de regressão aleatória usando como bases as funções polinomiais de Legendre, de Jacobi modificadas e trigonométricas, com aplicação na análise genética dos pesos de bovinos da raça Nelore.** 2007. 101 f. Tese (Doutorado em Agronomia, Área de concentração: Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2007.

MAGNABOSCO, C.D.U.; FARIA, C.U.; BORJAS, A.L.R.; LÔBO, R.R.; SAINZ, R.D. **Implementação da amostragem de Gibbs para a estimação de componentes de co-variância e parâmetros genéticos em dados de campo de bovino Nelore.** Planaltina: Embrapa/CPAC, 2001. 48 f. (Documentos, Embrapa/CPAC, 37).

MAGNABOSCO, C.D.U.; LÔBO, R.B.; FAMULA, T.R. Bayesian inference for genetic parameter estimation on growth traits for Nelore cattle in Brazil, using the Gibbs sampler. **Journal animal breeding and genetics**, v.117, p.169-188. 2000.

MAO, J.; ZHU, Z.; FUNG, W.K. Joint estimation of mean-covariance model for longitudinal data with basis function approximations. **Computational Statistics and Data Analysis**. V.55,p.983-992. 2011.

MASCIOLI, A.S.; ALENCAR, M.M. FREITAS, A.R.; MARTINS, E.N. Estudo da interação genótipo x ambiente sobre características de crescimento de bovinos de corte utilizando-se inferência bayesiana. **Revista Brasileira de Zootecnia**, v.35, n.6, p.2275-2284, 2006.

MELLO, S.P.; ALENCAR, M.M.; TORAL, F.L.B.; GIANLORENÇO, K. Estimativas de parâmetros genéticos para características de crescimento e produtividade em vacas da raça Canchim, utilizando-se inferência bayesiana. **Revista Brasileira de Zootecnia**, v.35, n.1, 2006. p.92-97.

MENDES, P. N. **Curvas de crescimento difásicas de fêmeas Hereford com erros autoregressivos e heterogeneidade de variâncias**. 2007. 98 f. Dissertação (Mestrado em Agronomia, Área de concentração: Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2007.

MISZTAL, I. Properties of random regression models using linear splines. **Journal Animal Breeding and Genetics**, v.123, p. 74-80, 2006.

MUCARI, T.B.; OLIVEIRA, J.A. Análise genético-quantitativa de pesos aos 8, 12, 18 e 24 meses de idade em um rebanho Guzerá. **Revista Brasileira de Zootecnia**, v.32, n.6, p.1604-1616. 2003 (Supl. 1).

NOGUEIRA, D.A.; SÁFADI, T.; BEARZOTI, E.; BUENO FILHO, J.S.S. Análises clássica e bayesiana de um modelo misto aplicado ao melhoramento animal: uma ilustração. **Ciência e Agrotecnologia**, edição especial, p.1614-1624, 2003.

NÚÑEZ-ANTÓN, V.; ZIMMERMAN, D.L. Modeling nonstationary longitudinal data. **Biometrics**, v.56, n.3, p. 669-705. 2000.

OLIVEIRA, H.N. **Análise genético-quantitativa da curva de crescimento de fêmeas da raça Guzerá**. 1995. 73 p. Tese (Doutorado em Ciências) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 1995.

PALA, A.; SAVAS, T.; UGUR, F.; DAS, G. Growth curves of Turkish Saanem goats' grouped for weight and body mass index. **Arch. Tierz.**, v.48, n.2, p.185-193. 2005.

PARK, T.; CASELLA, G. The Bayesian Lasso. **Journal of the American Statistical Association**. v103, n.482, p.681-686. 2008.

PLETCHER, S.D.; GEYER, C.J. The genetic analysis of age-dependent traits: modeling the character process. **Genetics**, v.151, p.825-835. 1999.

POURAHMADI, M. Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. **Biometrika**, v.86, p.677-690. 1999.

RESENDE, M.D.V. **Inferência bayesiana e simulação estocástica (Amostragem de Gibbs) na estimação de componentes de variância e de valores genéticos em plantas perenes**. Colombo: Embrapa/CNPQ, 2000. 68 f. (Embrapa/CNPQ, Documentos, 46).

RESENDE, M.D.V.; DUDA, L.L.; GUIMARÃES, P.R.B.; FERNANDES, J.S.C. Análise de modelos lineares mistos via inferência bayesiana. **Revista de Matemática e Estatística**, v.19, p.41-70, 2001.

SARMENTO, J.L.R.; PIMENTA FILHO, E.C.; RIBEIRO, M.N.; MARTINS FILHO, R. Efeitos ambientais e genéticos sobre o ganho em peso diário de bovinos Nelore no estado da Paraíba. **Revista Brasileira de Zootecnia**, v.32, n.2, p.325-330. 2003.

SIQUEIRA, R.L.P.G.; OLIVEIRA, J.A.; LÔBO, R.B.; BEZERRA, L.A.F.; TONHATI, H. Análise da variabilidade genética aditiva de características de crescimento na raça Nelore. **Revista Brasileira de Zootecnia**, v.32, n.1., p.99-105. 2003.

TIAO, G.C.; ZELLNER, A. On the bayesian estimation of multivariate regression. **Journal of the Royal Statistical Society**, v.26, n.2, p.277-285. 1964.

TORAL, F.L.B.; ALENCAR, M.M.; FREITAS, A.R. Abordagens freqüentista e bayesiana para avaliação genética de bovinos da raça Canchim para características de crescimento. **Revista Brasileira de Zootecnia**, v.36, n.1. 2007. p.43-53.

TORRES, J.M.C.; SANCHES, A.; CROSSA, A. Abordagem bayesiana de modelos para análise da interação genótipo por ambiente no melhoramento de plantas. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 52, 2007. Santa Maria. **Anais...** Santa Maria: RBRAS. 2007. [CD-ROM]

VAN TASSEL, C.P.; VAN VLECK, L.D. Multiple-trait Gibbs Sampler for animal models: flexible programs for bayesian and likelihood-based (co)variance component inference. **Journal Animal Science**, v.74, p.2586-2597. 1996.

WAAGEPETERSEN, R. INANEZ-ESCRICHE, N. SORENSEN, D. A comparison of strategies for Markov chain Monte Carlo computation in quantitative genetics. **Geneti. Sel. Evol**, v.40, p.161-176. 2008.

WALDMANN,P.; HALLANDER, J.; HOTI, F.; SILLANPAA, M.J. Efficient markov chain monte carlo implementation of Bayesian analysis of additive and dominance genetic variances in noinbred pedigrees.

WINTER, E.M.W.; ALMEIDA, M.I.M.; OLIVEIRA, E.G.; MARTINS, E.N.; NATEL, A.S.; SUREK, D. Aplicação do método Bayesiano na estimação de correlações genéticas e fenotípicas de peso em codornas de corte e várias idades. **Revista Brasileira de Zootecnia**, v.35, n.4 (supl.), p.1684-1690, 2006.

WOLFINGER, R.D. Heterogeneous variance: covariance structures for repeated measures. **Journal of Agriculture, Biological and Environmental Statistics**, v.1, n.2, p.205-230. 1996.

WU, W.B.; POURAHMADI, M. Nonparametric estimation of large covariance matrices of longitudinal data. **Biometrika**, v.90, p.831-844. 2003.

YAMAKI, M. **Estimação de parâmetros genéticos de produção de leite e de gordura da raça Pardo-Suiça, utilizando metodologias freqüentistas e bayesiana.** 2006. 71 f. Dissertação (Mestrado em Zootecnia) – Universidade Federal de Viçosa, Viçosa, 2006.

ZIMMERMAN, D.L.; NÚÑEZ-ANTÓN, V.; EL-BARMI, H. Computational aspects of likelihood-based estimation of first-order antedependence models. *Journal of Statitics, **Computation and Simulation***, v.60, p.67-84. 1998.

APÊNDICE

CODIGO FONTE DO PROGRAMA ADBAYES

PROGRAM MODELAP

```
!-----
USE PROB; USE DENSEOP; USE RANLIB; USE KINDS; USE LAPACK90; USE SPARSEM
!-----
```

IMPLICIT NONE

```
TYPE(SPARSE_HASHM), POINTER::AINV_H(:), XX_H(:)
TYPE(SPARSE_IJA), POINTER::AINV_IJA(:), XX_IJA(:)
REAL, ALLOCATABLE:: XY(:, :), VARAP(:, :), VARG(:, :), MIV(:), GIBBSA(:, :), Sap(:), Sap2(:) &
, SAG(:), SAG2(:), IV_G(:), SOLT(:), VART(:, :), IV_A(:), CHUTE(:) &
, MATRIZVG(:, :), MATRIZVP(:, :), AUGT(:), AUAT(:), AU_G(:, :), REGC1(:) &
, AU_A(:, :), AJUST(:), SVV_G(:, :), SVY_G(:), SVAV_G(:, :), SVAY_G(:) &
, SVV_P(:, :), SVY_P(:), SVAV_P(:, :), SVAY_P(:), EREr(:), REGC2(:) &
, VPRG(:, :), VPRA(:, :))
REAL*8, ALLOCATABLE:: SOL(:), Ci(:, :))
REAL (R8), ALLOCATABLE:: VARef(:, :), MEDef(:), AJUSB(:), XYv(:), AJUSG(:), VRES(:), XYTB(:) &
, LVA_G(:, :), LMA_G(:), LVA_P(:, :), LMA_P(:)
REAL (R8):: MEDau, VARau, LM, LV, VARefS, MEDefS, GEN(3)
REAL:: VT, SHAPE, SCALE, VMED, VGC, VAP, VG &
, SVV, SVY, SVAV, SVAY, FQ1, AJUS, CISC, DAT(5), XYT, SVVI
INTEGER:: I, G, BI, BF, NINTER, K, TAINV, J, IDEQ, NUMP, REG, NUMG, NEQ, GENT, APT, LOC, LOC2, L1 &
, L2, Z, PARENT, BIT, BFT, ADG, ADA, GC, NPAR, NPOLG, NPOLA, CR, I1, I2, F, CTL
INTEGER, ALLOCATABLE:: GPOL_G(:), GPOL_A(:), VLOC(:)
```

```
PRINT*,"*****"
PRINT*,"ADBAYES 1.0 - PROGRAMA PARA ANALISE DE ANTEDEPENDENCIA UTILIZANDO INFERENCIA"
PRINT*," BAYESIANA E METODOS DE MONTE CARLO E CADEIAS DE MARKOV "
PRINT*,"(VERSAO BETA 1.0)"
PRINT*,"AUTORES: FRANCISCO RIBEIRO DE ARAUJO NETO"
PRINT*," HENRIQUE NUNES DE OLIVEIRA"
PRINT*,"*****"
PRINT*," "
PRINT*," "
PRINT*,"-----"
PRINT*,"=====DEFINICAO DOS PARAMETROS GERAIS===== "
PRINT*,"-----"
PRINT*," NUMERO DE ANIMAIS DA MATRIZ DE PARENTESCO "
READ*, NUMG
PRINT*," NUMERO DE ANIMAIS COM REGISTRO "
READ*, NUMP
PRINT*," NUMERO DE MEDIDAS "
READ*, REG
PRINT*," NUMERO DE NIVEIS DO EFEITO FIXO-1 "
READ*, GC
```

GEN=NUMG*REG; APT=NUMP*REG; NEQ=REG+GC+GENT+APT

```
ALLOCATE(XX_H(REG), XX_IJA(REG), AINV_H(1), AINV_IJA(1), XY(NEQ, REG), SOL(NEQ), AJUSB(REG), XYv(REG) &
, IV_G(REG), VARG(REG, REG), VARAP(REG, REG), AJUSG(REG), SOLT(REG), VART(REG, REG), SAG(GENT) &
, SAG2(GENT), SAP(APT), SAP2(APT), VAREF(REG, REG), MEDEF(REG), MIV(REG), Ci(REG, REG) &
, IV_A(REG), VLOC(GENT), VRES(REG), XYTB(REG), AJUST(REG), EREr(REG))
```

```
PRINT*," "
PRINT*," "
PRINT*,"-----"
PRINT*,"=====DEFINICAO DOS PARAMETROS DE ANTEDEPENDENCIA===== "
PRINT*,"-----"
PRINT*," "
PRINT*,"-----EFEITO GENETICO ADITIVO-----"
PRINT*," "
PRINT*," "
PRINT*," "
PRINT*," ORDEM DE ANTEDEPENDENCIA "
READ*, ADG
```

ALLOCATE(GPOL_G(ADG))

```
PRINT*," ORDEM DOS POLINOMIOS A SEREM ALLOCADOS "
DO I=1, ADG; PRINT*, "ORDEM", I; READ*, GPOL_G(I); END DO
PRINT*," "
PRINT*,"-----EFEITO AMBIENTE PERMANENTE-----"
PRINT*," "
PRINT*," "
PRINT*," ORDEM DE ANTEDEPENDENCIA "
PRINT*," "
```

```

READ*,ADA

      ALLOCATE(GPOL_A(ADA))

PRINT*,"          ORDEM DOS POLINOMIOS A SEREM ALLOCADOS          "
      DO I=1,ADA; PRINT*,"ORDEM",I;READ*,GPOL_A(I);END DO
PRINT*,"          "
PRINT*,"-----"
PRINT*,"=====DEFINICAO DOS PARAMETROS DE ARQUIVOS=====
PRINT*,"-----"
PRINT*,"          NUMERO DE AMOSTRAS          "
READ*,NINTER
PRINT*,"          ARQUIVO DE INVERSA"
READ*,TAINV

NPOLG=SUM(GPOL_G);NPOLA=SUM(GPOL_A)

NPAR=NPOLG+REG+NPOLA+REG+REG

ALLOCATE(GIBBSA(NINTER,NPAR),CHUTE(NPAR),AUGT(NPOLG),AUAT(NPOLA),MATRIZVG(GENT,NPOLG),MATRIZVP(APT,NPOLA)
&
      ,AU_G(ADG,GPOL_G(1)),AU_A(ADA,GPOL_A(1)),SVV_G(NPOLG,NPOLG),SVY_G(NPOLG),SVAV_G(NPOLG,NPOLG)  &
      ,SVAY_G(NPOLG),SVV_P(NPOLA,NPOLA),SVY_P(NPOLA),SVAV_P(NPOLA,NPOLA),SVAY_P(NPOLA)          &
      ,LVA_G(NPOLG,NPOLG),LMA_G(NPOLG),LVA_P(NPOLA,NPOLA),LMA_P(NPOLA),REGC1(NPOLG),REGC2(NPOLA)  &
      ,VPRG(NPOLG,NPOLG),VPRA(NPOLA,NPOLA))

PRINT*,"-----"
PRINT*,"=====DEFINICAO DOS CHUTES=====
PRINT*,"-----"
PRINT*,"          "
PRINT*,"-----GENETICO ADITIVO-----"
PRINT*,"          REGRESSORES-AUTORREGRESSIVOS"
READ*,CHUTE(ANDPAR(1,0):ANDPAR(1,1))
PRINT*,"          VARIANCIAS DAS INOVACOES"
READ*,CHUTE(ANDPAR(2,0):ANDPAR(2,1))
PRINT*,"
PRINT*,"-----AMBIENTE PERMANENTE - ANIMAL-----"
PRINT*,"          REGRESSORES-AUTORREGRESSIVOS"
READ*,CHUTE(ANDPAR(3,0):ANDPAR(3,1))
PRINT*,"          VARIANCIAS DAS INOVACOES"
READ*,CHUTE(ANDPAR(4,0):ANDPAR(4,1))
PRINT*,"
PRINT*,"-----VARIANCIA RESIDUAL-CLASSES-----"
READ*,CHUTE(ANDPAR(5,0):ANDPAR(5,1))

PRINT*,"          "
PRINT*,"          "

SOL=0.D0;GIBBSA=0.D0;XY=0.D0
VPRG=0.D0;VPRA=0.D0
DO I=1,NPOLG; VPRG(I,I)=1/10000;END DO;
DO I=1,NPOLA; VPRA(I,I)=1/10000;END DO;

!*****
!          FORMACAO DA MATRIX XX E VETOR XY
!*****

DO I=1,REG
CALL ZEROM(XX_H(I),NEQ);CALL ZEROM(XX_IJA(I),NEQ)
END DO

      !-----!
      !ESTRUTURA DO ARQUIVO DE DADOS!
      !1-REGISTROS          !
      !2-INDIVIDUO GENETICO      !
      !3-INDIVIDUO AMBIENTE      !
      !4-TEMPO          !
      !5-GRUPOS DE CONTEMPORANEOS !
      !-----!

OPEN(UNIT=200,FILE='dados.txt',STATUS='OLD')
DO I=1,APT
      READ(200,*) DAT(1:5)

!MEDIA

```

```

CALL ADD_SPARSE_HASHM(1.D0,ANDRESS(INT(DAT(4)),1),ANDRESS(INT(DAT(4)),1),XX_H(INT(DAT(4))))
  XY(ANDRESS(INT(DAT(4)),1),INT(DAT(4)))=XY(ANDRESS(INT(DAT(4)),1),INT(DAT(4)))+DAT(1)

!EFEITO FIXO EXTRA
CALL ADD_SPARSE_HASHM(1.D0,ANDRESS(INT(DAT(5)),2),ANDRESS(INT(DAT(5)),2),XX_H(INT(DAT(4))))
  XY(ANDRESS(INT(DAT(5)),2),INT(DAT(4)))=XY(ANDRESS(INT(DAT(5)),2),INT(DAT(4)))+DAT(1)

!AMBIENTE PERMANENTE
CALL ADD_SPARSE_HASHM(1.D0,ANDRESS(INT(DAT(3)),3,INT(DAT(4))),ANDRESS(INT(DAT(3)),3,INT(DAT(4))),XX_H(INT(DAT(4))))
  XY(ANDRESS(INT(DAT(3)),3,INT(DAT(4))),INT(DAT(4)))=XY(ANDRESS(INT(DAT(3)),3,INT(DAT(4))),INT(DAT(4)))+DAT(1)

!GENETICO ADITIVO
CALL ADD_SPARSE_HASHM(1.D0,ANDRESS(INT(DAT(2)),4,INT(DAT(4))),ANDRESS(INT(DAT(2)),4,INT(DAT(4))),XX_H(INT(DAT(4))))
  XY(ANDRESS(INT(DAT(2)),4,INT(DAT(4))),INT(DAT(4)))=XY(ANDRESS(INT(DAT(2)),4,INT(DAT(4))),INT(DAT(4)))+DAT(1)
  VLOC(INT(DAT(2)))=1

!EFEITO FIXO/EFEITO ALEATORIO
CALL ADD_SPARSE_HASHM(1.D0,ANDRESS(INT(DAT(4)),1),ANDRESS(INT(DAT(5)),2),XX_H(INT(DAT(4))))
CALL ADD_SPARSE_HASHM(1.D0,ANDRESS(INT(DAT(4)),1),ANDRESS(INT(DAT(3)),3,INT(DAT(4))),XX_H(INT(DAT(4))))
CALL ADD_SPARSE_HASHM(1.D0,ANDRESS(INT(DAT(4)),1),ANDRESS(INT(DAT(2)),4,INT(DAT(4))),XX_H(INT(DAT(4))))

CALL ADD_SPARSE_HASHM(1.D0,ANDRESS(INT(DAT(5)),2),ANDRESS(INT(DAT(3)),3,INT(DAT(4))),XX_H(INT(DAT(4))))
CALL ADD_SPARSE_HASHM(1.D0,ANDRESS(INT(DAT(5)),2),ANDRESS(INT(DAT(2)),4,INT(DAT(4))),XX_H(INT(DAT(4))))

CALL ADD_SPARSE_HASHM(1.D0,ANDRESS(INT(DAT(3)),3,INT(DAT(4))),ANDRESS(INT(DAT(2)),4,INT(DAT(4))),XX_H(INT(DAT(4))))

ENDDO
CLOSE(200)

DO I=1,REG
CALL CONVERT_HASH_IJA_GENERAL(XX_IJA(I),XX_H(I),CONV_UPPER_FULL)
CALL RESET(XX_H(I))
END DO

35 FORMAT(100(1X,ES13.7))

OPEN(UNIT=201,FILE='samples.txt',STATUS='NEW')

!*****
!          FORMACAO DA INVERSA DA MATRIZ DE PARENTESCO
!*****

DO I=1,1
CALL ZEROM(AINV_H(I),NUMG);CALL ZEROM(AINV_IJA(I),NUMG)
END DO

OPEN(UNIT=202,FILE='Ainv.txt',STATUS='OLD')
DO I=1,TAINV
  READ(202,*) GEN(1:3)
CALL ADD_SPARSE_HASHM(GEN(3),INT(GEN(2)),INT(GEN(1)),AINV_H(1))
ENDDO
CLOSE(202)
CALL CONVERT_HASH_IJA_GENERAL(AINV_IJA(1),AINV_H(1),CONV_UPPER_FULL)

CALL RESET(AINV_H(1))

!*****
!          CADEIA DE GIBBS
!*****

DO G=1,NINTER

IF (G==1) THEN

  CTL=0;AUGT=0.D0
  AUGT=CHUTE(ANDPAR(1,0):ANDPAR(1,1))
  DO J=1,ADG
    DO I=1,GPOL_G(J)
      AU_G(J,I)=AUGT(I+CTL);
    END DO
    CTL=CTL+GPOL_G(J)
  END DO

  CTL=0;AUAT=0.D0

```

```

      AUAT=CHUTE(ANDPAR(3,0):ANDPAR(3,1))
DO J=1,ADA
  DO I=1,GPOL_A(J)
    AU_A(J,I)=AUAT(I+CTL);
  END DO
  CTL=CTL+GPOL_A(J)
END DO

IV_G=CHUTE(ANDPAR(2,0):ANDPAR(2,1))
IV_A=CHUTE(ANDPAR(4,0):ANDPAR(4,1))
VRES=CHUTE(ANDPAR(5,0):ANDPAR(5,1))
ELSE

  CTL=0.D0;AUGT=0.D0
  AUGT=GIBBSA(G-1,ANDPAR(1,0):ANDPAR(1,1))
DO J=1,ADG
  DO I=1,GPOL_G(J)
    AU_G(J,I)=AUGT(I+CTL);
  END DO
  CTL=CTL+GPOL_G(J)
END DO

  CTL=0.D0;AUAT=0.D0
  AUAT=GIBBSA(G-1,ANDPAR(3,0):ANDPAR(3,1))
DO J=1,ADA
  DO I=1,GPOL_A(J)
    AU_A(J,I)=AUAT(I+CTL);
  END DO
  CTL=CTL+GPOL_A(J)
END DO

IV_G=GIBBSA(G-1,ANDPAR(2,0):ANDPAR(2,1))
IV_A=GIBBSA(G-1,ANDPAR(4,0):ANDPAR(4,1))

VRES=GIBBSA(G-1,ANDPAR(5,0):ANDPAR(5,1))

END IF

CALL VARAD3(AU_G,IV_G,VARG,GPOL_G) !OBTEN A INVERSA DA MATRIZ DE VARIANCIA GENETICA
CALL VARAD3(AU_A,IV_A,VARAP,GPOL_A) !OBTEN A INVERSA DA MATRIZ DE VARIANCIA AMBIENTE PERMANENTE

!*****
!
!          POSTERIORI PARA OS EFEITOS FIXOS E ALEATORIOS
!*****
!-----
!AMOSTRAGEM PARA OS EFEITOS FIXOS: GIBBS-ESCALAR
!-----
DO I=1,REG+GC

AJUS=0.D0;CISC=0.D0;XYT=0.D0

DO CR=1,REG
  DO K=1,XX_IJA(CR)%IA(I+1)-XX_IJA(CR)%IA(I)
    LOC=INT(XX_IJA(CR)%IA(I))+K-1
    IF (XX_IJA(CR)%JA(LOC)==I) THEN
      CISC=CISC+XX_IJA(CR)%A(LOC)*(1.D0/VRES(CR))
    ELSE
      AJUS=AJUS+XX_IJA(CR)%A(LOC)*(1.D0/VRES(CR))*SOL(INT(XX_IJA(CR)%JA(LOC)))
    END IF
  END DO
  XYT=XYT+XY(I,CR)*(1.D0/VRES(CR))
END DO

LV=1.D0/CISC
LM=LV*(XYT-AJUS)
SOL(I)=GEN_NORMAL_SCALAR(LM,LV)

END DO
!-----
!AMOSTRAGEM PARA OS EFEITOS DE AMBIENTE PERMANENTE: GIBBS-BLOCO
!-----
DO I=1,NUMP

Ci=0.D0;AJUSB=0.D0;VARRef=0.D0;MEDef=0.D0;AJUSG=0.D0;XYTB=0.d0

```

```

BI=ANDRESS(I,3,1);BF=ANDRESS(I,3,REG)

!CORRECAO PARA EFEITOS NAO-CORRELACIONADOS
DO F=1,REG
  DO J=1,REG
    AJUS=0.D0
    IDEQ=ANDRESS(I,3,J)
    DO K=1,XX_IJA(F)%IA(IDEQ+1)-XX_IJA(F)%IA(IDEQ)
      LOC=INT(XX_IJA(F)%IA(IDEQ))+K-1
      IF (XX_IJA(F)%JA(LOC)==IDEQ) THEN
        AJUS=AJUS+0.D0
      ELSE
        AJUS=AJUS+XX_IJA(F)%A(LOC)*(1.D0/VRES(F))*SOL(INT(XX_IJA(F)%JA(LOC)))
      END IF
    END DO
    AJUSB(J)=AJUSB(J)+AJUS
  END DO
  XYTB=XYTB+XY(BI:BF,F)*(1.D0/VRES(F))
END DO

DO K=1,REG; Ci(K,K)=(1.D0/VRES(K));END DO
Ci=Ci+VARAP

CALL INVERSE(CI);
VAREf=Ci
AJUST=XYTB-AJUSB
MEDef=MATMUL(VAREf,AJUST)
SOL(BI:BF)=GEN_NORMAL_MULT(I(MEDef,VAREf)

END DO
!-----
!AMOSTRAGEM PARA OS EFEITOS GENETICOS ADITIVOS: GIBBS-BLOCO
!-----
DO I=1,NUMG

Ci=0.D0;AJUSB=0.D0;VAREf=0.D0;MEDef=0.D0;AJUSG=0.D0;XYTB=0.d0
BI=ANDRESS(I,4,1);BF=ANDRESS(I,4,REG)

!CORRECAO PARA EFEITOS NAO-CORRELACIONADOS
DO F=1,REG
  DO J=1,REG
    AJUS=0.D0
    IDEQ=ANDRESS(I,4,J)
    DO K=1,XX_IJA(F)%IA(IDEQ+1)-XX_IJA(F)%IA(IDEQ)
      LOC=INT(XX_IJA(F)%IA(IDEQ))+K-1
      IF (XX_IJA(F)%JA(LOC)==IDEQ) THEN
        AJUS=AJUS+0.D0
      ELSE
        AJUS=AJUS+XX_IJA(F)%A(LOC)*(1.D0/VRES(F))*SOL(INT(XX_IJA(F)%JA(LOC)))
      END IF
    END DO
    AJUSB(J)=AJUSB(J)+AJUS
  END DO
  XYTB=XYTB+XY(BI:BF,F)*(1.D0/VRES(F))
END DO

!CONSTRUCAO DA MATRIZ BASE PARA AS VARIANCIAS
IF(VLOC(I)==1) THEN
  DO K=1,REG
    CI(K,K)=1.D0/VRES(K)
  END DO
END IF

!CORRECAO PARA OS EFEITOS ADITIVOS
AJUSG=0.D0

L1=INT(AINV_IJA(1)%IA(I+1))-INT(AINV_IJA(1)%IA(I))
DO Z=1,L1
  L2=INT(AINV_IJA(1)%IA(I))+Z-1
  PARENT=AINV_IJA(1)%JA(L2)
  IF(PARENT==I) THEN
    CI=CI+VARG*AINV_IJA(1)%A(L2)
  ELSE
    BIT=ANDRESS(INT(PARENT),4,1);BFT=ANDRESS(INT(PARENT),4,REG)
    SOLT=SOL(BIT:BFT)
    VART=VARG*AINV_IJA(1)%A(L2)
    AJUSG=AJUSG+MATMUL(VART,SOLT)
  END IF
END DO

```

```

        END IF
    END DO

    CALL INVERSE(CI);
    VARef=Ci
    AJUST=XYTB-AJUSB-AJUSG
    MEDef=MATMUL(VARef,AJUST)
    SOL(BI:BF)=GEN_NORMAL_MULT1(MEDef,VARef)

    END DO

!*****
!          POSTERIOR PARA VARIANCIA GENETICA ADITIVA
!*****

SAG=SOL(ANDRESS(1,4,1):ANDRESS(NUMG,4,REG))

!-----
!POSTERIOR PARA O PARAMETRO AUTORREGRESSIVO
!-----

SVAV_G=0.D0; SVAY_G=0.D0

CALL VMATRIZ2(NUMG,REG,SAG,GPOL_G,MATRIZVG,1)

DO I=1,NUMG
  DO K=1,REG
    LOC=REG*(I-1)+K
    IF (K==1) THEN
      SVAV_G=SVAV_G+0.D0; SVAY_G=SVAY_G+0.D0
    ELSE
      L1=INT(AINV_IJA(1)%IA(I+1))-INT(AINV_IJA(1)%IA(I))
      SVV_G=0.D0; SVY_G=0.D0
      DO Z=1,L1
        L2=INT(AINV_IJA(1)%IA(I))+Z-1
        PARENT=AINV_IJA(1)%JA(L2)
        LOC2=REG*(PARENT-1)+K
        DO I1=1,NPOLG
          SVY_G(I1)=SVY_G(I1)+MATRIZVG(LOC2,I1)*SAG(LOC)*(1.D0/IV_G(K))*AINV_IJA(1)%A(L2)
          DO I2=1,NPOLG
            SVV_G(I1,I2)=SVV_G(I1,I2)+MATRIZVG(LOC2,I1)*MATRIZVG(LOC,I2)*(1.D0/IV_G(K))*AINV_IJA(1)%A(L2)
          END DO
        END DO
      END DO
      SVAV_G=SVAV_G+SVV_G;SVAY_G=SVAY_G+SVY_G
    END IF
  END DO
END DO

SVAV_G=VPRG+SVAV_G
CALL INVERSE(SVAV_G)
LVA_G=SVAV_G; LMA_G=MATMUL(LVA_G,SVAY_G)
GIBBSA(G,ANDPAR(1,0):ANDPAR(1,1))=GEN_NORMAL_MULT1(LMA_G,LVA_G)
REGC1(1:NPOLG)=GIBBSA(G,ANDPAR(1,0):ANDPAR(1,1))

!-----
!POSTERIOR PARA AS INOVACOES DAS VARIANCIA
!-----

Sag2=0.D0

CALL VMATRIZ2(NUMG,REG,SAG,GPOL_G,MATRIZVG,2,REGC1,SAG2)

DO I=1,REG
  FQ1=0.D0
  DO K=1,NUMG
    LOC=REG*(K-1)+I
    SVVI=0.D0
    L1=INT(AINV_IJA(1)%IA(K+1))-INT(AINV_IJA(1)%IA(K))
    DO Z=1,L1
      L2=INT(AINV_IJA(1)%IA(K))+Z-1
      PARENT=AINV_IJA(1)%JA(L2)
      LOC2=REG*(PARENT-1)+I

```



```

                SVVI=SVVI+SAG2(LOC2)*SAG2(LOC)*AINV_IJA(1)%A(L2)
            END DO
        FQ1=FQ1+SVVI
        END DO

        Scale=FQ1/2.D0; Shape=NUMG/2.D0
        GIBBSA(G,I+ANDPAR(1,1))=1.D0/GENGAM(scale,shape)
        END DO

!*****
!          POSTERIOR PARA A VARIANCIA DE AMBIENTE PERMANENTE
!*****

        Sap=SOL(ANDRESS(1,3,1):ANDRESS(NUMP,3,REG))

!-----
!POSTERIOR PARA O PARAMETRO AUTORREGRESSIVO
!-----
        SVAV_P=0.D0; SVAY_P=0.D0

        CALL VMATRIZ2(NUMP,REG,SAP,GPOL_A,MATRIZVP,1)

        DO I=1,NUMP
            DO K=1,REG
                LOC=REG*(I-1)+K
                IF (K==1) THEN
                    SVAV_P=SVAV_P+0.D0; SVAY_P=SVAY_P+0.D0
                ELSE
                    DO I1=1,NPOLA
                        SVY_P(I1)=SVY_P(I1)+MATRIZVP(LOC,I1)*SAP(LOC)*(1.D0/IV_A(K))
                        DO I2=1,NPOLA
                            SVV_P(I1,I2)=SVV_P(I1,I2)+MATRIZVP(LOC,I1)*MATRIZVP(LOC,I2)*(1.D0/IV_A(K))
                        END DO
                    END DO
                    SVAV_P=SVAV_P+SVV_P;SVAY_P=SVAY_P+SVY_P
                END IF
            END DO
        END DO

        SVAV_P=VPRA+SVAV_P
        CALL INVERSE(SVAV_P)
        LVA_P=SVAV_P; LMA_P=MATMUL(LVA_P,SVAY_P)
        GIBBSA(G,ANDPAR(3,0):ANDPAR(3,1))=GEN_NORMAL_MULTI(LMA_P,LVA_P)
        REGC2(1:NPOLA)=GIBBSA(G,ANDPAR(3,0):ANDPAR(3,1))

!-----
!POSTERIOR PARA AS INOVACOES DAS VARIANCIA
!-----

        SAP2=0.D0

        CALL VMATRIZ2(NUMP,REG,SAP,GPOL_A,MATRIZVP,2,REGC2,SAP2)

        DO I=1,REG
            FQ1=0.D0
            DO K=1,NUMP
                LOC=REG*(K-1)+I
                FQ1=FQ1+SAP2(LOC)*SAP2(LOC)
            END DO
            Scale=FQ1/2.D0; Shape=NUMP/2.D0
            GIBBSA(G,I+ANDPAR(3,1))=1.D0/GENGAM(scale,shape)
        END DO

!*****
!          POSTERIORI PARA A VARIANCIA RESIDUAL
!*****

        erer=0.D0

        OPEN(UNIT=200,FILE='dados.txt',STATUS='OLD')
        DO I=1,APT
            READ(200,*) DAT(1:5)
            VMED=SOL(ANDRESS(INT(DAT(4)),1));VGC=SOL(ANDRESS(INT(DAT(5)),2))
            VAP=SOL(ANDRESS(INT(DAT(3)),3,INT(DAT(4)))));VG=SOL(ANDRESS(INT(DAT(2)),4,INT(DAT(4))))
            erer(INT(DAT(4)))=erer(INT(DAT(4)))+(DAT(1)-VMED-VGC-VAP-VG)**2.D0
        END DO

```

```

ENDDO
CLOSE(200)

DO I=1,REG
Scale=ERER(I)/2.d0
Shape=NUMP/2.d0
GIBBSA(G,I+ANDPAR(4,1))=1/GENGAM(scale,shape)
END DO

```

```

!=====
!RESULTADOS CICLO
!=====
PRINT*,"*****"
PRINT*,"*****"
PRINT*,"CICLO",G
PRINT*,"*****"
PRINT*," "
PRINT*,"-----VARIANCIA GENETICA-----"
PRINT*,"REGRESSORES"
PRINT*,GIBBSA(G,ANDPAR(1,0):ANDPAR(1,1))
PRINT*,"VARIANCIAS DAS INOVACOES"
PRINT*,GIBBSA(G,ANDPAR(2,0):ANDPAR(2,1))
PRINT*," "
PRINT*,"-----VARIANCIA AMBIENTE-----"
PRINT*,"REGRESSORES"
PRINT*,GIBBSA(G,ANDPAR(3,0):ANDPAR(3,1))
PRINT*,"VARIANCIAS DAS INOVACOES"
PRINT*,GIBBSA(G,ANDPAR(4,0):ANDPAR(4,1))
PRINT*," "
PRINT*,"-----VARIANCIA RESIDUAL-----"
PRINT*,GIBBSA(G,ANDPAR(5,0):ANDPAR(5,1))
PRINT*,"*****"
!=====

```

```
WRITE(201,35) GIBBSA(G,1:NPAR)
```

```
END DO
```

```
END PROGRAM MODELAP
```