

LAIS COSTA BRITO

**REGRESSÃO ALEATÓRIA NA AVALIAÇÃO DA PRODUÇÃO DE LEITE DE  
BÚFALAS MURRAH UTILIZANDO INFERÊNCIA BAYESIANA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-graduação em Zootecnia, para obtenção do título de *Magister Scientiae*.

VIÇOSA  
MINAS GERAIS – BRASIL  
2013

**Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV**

T

B862r  
2013

Brito, Lais Costa, 1990-  
Regressão aleatória na avaliação da produção de  
leite de búfalas Murrah utilizando inferência bayesiana / Lais  
Costa Brito. – Viçosa, MG, 2013.  
x, 47 f. : il. ; 29 cm.

Orientador: Ricardo Frederico Euclides.  
Dissertação (mestrado) - Universidade Federal de Viçosa.  
Referências bibliográficas: f. 40-47.

1. Búfalo - Melhoramento genético. 2. Leite - Produção -  
Avaliação. 3. Polinômios. 4. Teoria bayesiana de decisão  
estatística. I. Universidade Federal de Viçosa. Departamento de  
Zootecnia. Programa de Pós-Graduação em Zootecnia.  
II. Título.

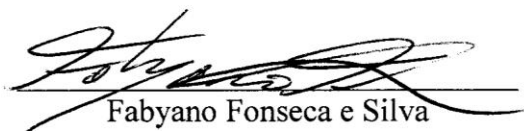
CDD 22. ed. 636.2930821

LAIS COSTA BRITO


**REGRESSÃO ALEATÓRIA NA AVALIAÇÃO DA PRODUÇÃO DE LEITE DE  
BÚFALAS MURRAH UTILIZANDO INFERÊNCIA BAYESIANA**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-graduação em Zootecnia, para obtenção do título de *Magister Scientiae*.

APROVADA: 19 de março de 2013.

  
Fabyano Fonseca e Silva

  
Paulo Sávio Lopes

  
Robledo de Almeida Torres  
(Presidente da Banca)

*Aos meus pais, irmãs e amigos,*

*Que a cada dia me ensinam a viver da melhor forma possível;  
que me ensinaram a contradição que é ser feliz!*

*Dedico este trabalho.*

*“[...] A seleção natural diariamente, hora a hora faz um escrutínio no mundo todo das menores variações, rejeitando as más, preservando e complementando as boas, trabalhando silenciosa e insensivelmente [...]. Não vemos essas mudanças lentas em progresso, até que a mão do tempo coloque a marca de sua passagem, e ainda assim nossa visão das eras geológicas passadas é tão imperfeita, que vemos apenas que as formas de vida são agora diferentes do que foram [...]”*

*Darwin, The origin of species.*

## AGRADECIMENTOS

*“Aqueles que passam por nós, não vão sós, não nos deixam sós.*

*Deixam um pouco de si, levam um pouco de nós.”*

*Antoine de Saint-Exupéry*

Por isto só tenho a agradecer:

À Universidade Federal de Viçosa, em especial ao Departamento e ao Programa de Pós-graduação em Zootecnia, pela oportunidade de realização deste curso.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela concessão da bolsa de estudo.

Ao meu orientador, Prof. Ricardo Frederico Euclides (Bajá), que me fez perceber que os momentos mais simples são os mais importantes em nossas vidas.

Aos meus coorientadores, Prof. Robledo de Almeida Torres e Prof. Humberto Tonhati, pelo engrandecimento pessoal e profissional.

Ao grupo de criadores de búfalos associados à ABCB, pela concessão de dados que permitiram a realização deste trabalho.

Aos Prof. Paulo Sávio Lopes & Prof. Fabyano Fonseca e Silva, pela disponibilidade em participar da banca examinadora, sugestões que contribuíram para aprimorar este trabalho e apoio imprescindível nos momentos finais da conclusão desta etapa.

Aos amigos Francisco Araújo Neto e Raul Aspilcueta-Borquis, por não terem medido esforços para a concretização desta jornada, obrigada por tudo!

Aos meus pais, Cidia Costa e Miguel Brito, por terem investido incansavelmente na minha formação e por todo o chamego.

Às minhas irmãs, Beatriz Brito e Thais Brito, pelo eterno companheirismo. Nossa tríade é a força que me move diariamente.

Ao Douglas Nogueira (Paullista), simplesmente por tudo. Sem seu carinho e apoio não teria alcançado esse sonho. Serei eternamente grata.

À toda minha família pelo apoio incansável, em especial Mario Barata e minha avó Maria José.

A todos os meus mestres pelo ensinamento e preparo profissional, principalmente Prof. Cláudio Araújo e Luciara Chaves, os “culpados” que me conduziram ao mundo do melhoramento genético animal.

A todos os meus amigos de república e da vida. Em especial Cléo, Michael, Iara, Jerusa, Suellen, David e Emerson; por todos os momentos de alegria, apoio e paciência.

Aos amigos do Programa de Seleção Genética (PROSEG-UFRA/Belém-PA) - em nome da Prof<sup>ª</sup>. Luciara Chaves, Eduardo Daher e Prof<sup>º</sup>Alex Schierholt - por acreditarem em mim e por todo o ensinamento e paciência.

Aos amigos do Grupo de Melhoramento Genético de Caprinos (UFV): Ana Lúcia Melo, Felipe Silva, Geraldo de Souza Jr., Giovani Caetano, Hinayah Rojas, Joashlenny Oliveira, Luiz Brito, Nadson Oliveira, Paula Barros, Vinícius Junqueira, Yara Cunha e William Mariano (e demais que por descuido eu tenha esquecido). Sem vocês o trabalho não seguiria em frente!

Aos amigos da “Salinha” do Melhoramento, Labtec, da Genética e Zootecnia, pelas horas de estudo e companheirismo: Lucas, Renata, Luciano, Jeferson, Rodrigo “Batata”, Rodrigo, Janeo, Daniele, Aline, Helmut, Ana, Mariele, Carla, Evelyze, Elcer, Ali, Débora, Roman, Gabriela, David, Eric e tantos mais. Em especial Carolina, Edson, Felipe, Luiz, Matilde “Gansa”, Philipe “Novinho”, Thiago “Timão” e Vinícius. Vocês sempre estarão em minha memória e terão um lugar em meu coração.

Todas as pessoas que tive o prazer de conhecer em minha vida, pois foram imprescindíveis na consolidação do meu caráter.

A essa “força maior” que alguns chamam de Deus, outros de destino ou simplesmente coincidência. Na essência é o que nos move, nos dá esperança e nos faz perceber, dia-a-dia e em coisas simples, o quanto *“a vida é bonita e é bonita!”*.

## BIOGRAFIA

**LAIS COSTA BRITO**, solteira, filha de Miguel Baia Brito & Cidia Maria Rosário da Costa, nasceu em 26 de abril de 1990, em Belém, Pará. Em março de 2007, iniciou o curso de Zootecnia na Universidade Federal Rural da Amazônia – PA. No período de 2007/2008 foi estagiária no Centro de Pesquisa em Ovinos e Caprinos do Pará (CPCOP), com orientação do Prof. Luiz Fernando Rodrigues. Em outubro de 2008 recebeu o Diploma de Reconhecimento e Estímulo pelo seu destaque e aproveitamento curricular nas disciplinas do curso de Zootecnia. Entre os anos de 2008/2009 foi bolsista de Iniciação Científica da Embrapa Amazônia Oriental, com orientação do Dr. José Ribamar Felipe Marques. Entre os anos de 2010/2011 foi monitora da disciplina Melhoramento Genético Animal I, ministrada pela Prof. Luciara Celi Chaves. Em julho de 2011 recebeu o diploma de Zootecnista e em agosto do mesmo ano iniciou o curso de Mestrado em Zootecnia pela Universidade Federal de Viçosa. Em 19 de março de 2013, submeteu-se aos exames finais de defesa de dissertação para obtenção do título de *Magister Scientiae* em Zootecnia, sob orientação do Prof. Ricardo Euclides.

## SUMÁRIO

	Pág.
<b>RESUMO .....</b>	<b>vii</b>
<b>ABSTRACT .....</b>	<b>ix</b>
<b>1- INTRODUÇÃO.....</b>	<b>1</b>
<b>2- OBJETIVOS .....</b>	<b>3</b>
<b>3- REVISÃO BIBLIOGRÁFICA .....</b>	<b>4</b>
3.1. Histórico e situação da bubalinocultura .....	4
3.2. Avaliação de dados longitudinais .....	5
3.3. Polinômios segmentados .....	7
3.4. Inferência bayesiana.....	9
<b>4- METODOLOGIA.....</b>	<b>11</b>
4.1. Dados .....	11
4.2. Modelos.....	12
4.3. Comparação de modelos .....	15
<b>5- RESULTADOS E DISCUSSÃO.....</b>	<b>16</b>
5.1. Estrutura dos dados.....	16
5.2. Convergência e descrição da cadeia de Gibbs.....	17
5.2. Seleção de modelos.....	18
5.3. Descrição das estimativas dos parâmetros .....	20
5.4. Comparação dos valores genéticos (EBV): Produção de leite (PL270) e Persistência da lactação (PERSL) .....	29
<b>6- CONCLUSÕES.....</b>	<b>37</b>
<b>7- PERSPECTIVAS FUTURAS.....</b>	<b>38</b>
<b>8- REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>40</b>

## RESUMO

BRITO, Lais Costa, M.Sc., Universidade Federal de Viçosa, março de 2013. **Regressão Aleatória na Avaliação da Produção de Leite de Búfalas Murrah Utilizando Inferência Bayesiana.** Orientador: Ricardo Frederico Euclides. Coorientadores: Robledo de Almeida Torres e Humberto Tonhati.

Ao considerar a crescente importância da bubalinocultura na pecuária leiteira nacional, objetivou-se com o presente estudo modelar as variações da produção de leite no dia de controle (PLDC) de búfalas da raça Murrah por meio de modelos de regressão aleatória (polinômios ortogonais de Legendre – POL – e *B-splines* – BSP – lineares). Para tanto, foram analisados 15161 registros semanais de PLDC provenientes da primeira lactação de 1158 búfalas da raça Murrah, com partos registrados entre os anos de 1988 a 2005 e pertencentes a rebanhos do estado de São Paulo, Brasil. As PLDC foram divididas em 38 classes semanais e analisadas considerando modelo animal de regressão aleatória - que incluiu efeitos aleatórios genético-aditivo, de ambiente permanente e residual. Adicionalmente, incluíram-se efeitos fixos de grupo de contemporâneo, a covariável idade da búfala ao parto (efeito linear e quadrático), número de ordenhas e as médias dos coeficientes da curva de lactação representativa da população (sendo esta ajustada por meio de POL de 4ª ordem). As funções de covariâncias relativas aos coeficientes de regressão aleatória dos efeitos genético-aditivo e de ambiente permanente foram igualmente modeladas por meio de POL (3ª à 6ª ordem) ou BSP lineares (4 a 6 nós). A estrutura residual foi considerada heterogênea, contendo nove classes de variâncias. Os componentes de (co)variância dos parâmetros foram estimados via inferência bayesiana, utilizando o programa GIBBS3F90. A convergência das cadeias de Gibbs foi verificada por meio de inspeção gráfica e o pacote BOA, do programa R. As análises pós-gibbs foram feitas utilizando o POSTGIBBS1F90. A seleção dos melhores modelos se deu pela utilização do critério de informação da *deviance* (DIC). A média de PLDC foi igual a  $6,31 \pm 2,13$  kg. Os modelos POL de sexta ordem – POL6, e BSP com seis nós – BSP6 (8, 29, 57, 120, 239, 267 dias), com maior número de parâmetros (51) foram considerados os melhores para descrever a variação da PLDC ao longo da lactação. De modo geral, os parâmetros genéticos apresentaram tendências semelhantes para estes modelos, à exceção das estimativas de variância genética aditiva e herdabilidade, menores nas primeiras semanas de lactação para o modelo BSP6. Verificou-se que, apesar de pouco

correlacionados, os valores genéticos preditos para produção de leite aos 270 dias e persistência da lactação para os melhores modelos classificam os reprodutores de forma similar. O MRA mais recomendável para avaliação genética da PLDC de búfalos da raça Murrah é uma função *B-spline* (com seis nós), haja vista a procura por modelos mais parcimoniosos e com maiores vantagens computacionais para modelar variações da PLDC de búfalas da raça Murrah; e dado maior controle das estimativas de herdabilidade referentes aos extremos na curva de lactação.

## ABSTRACT

BRITO, Lais Costa, M.Sc., Universidade Federal de Viçosa, March, 2013. **Random Regression Analysis to Model Test-Day Milk Yield in Murrah Buffaloes using Bayesian Inference.** Adviser: Ricardo Frederico Euclides. Co-advisers: Robledo de Almeida Torres and Humberto Tonhati.

Considering the importance of raising buffaloes to the national dairy farming, the objective of this study was to model variations in test-day milk yield (TDMY) during the first lactation of Murrah buffaloes by random regression models (RRM), using orthogonal Legendre polynomials (POL) and linear B-splines (BSP). A total of 15,161 test-day records from 1,158 first lactations of Murrah buffaloes breed from herds belonging to São Paulo state, Brazil, with calving recorded from 1988 to 2005, were analyzed. For TDMY, 38 weekly classes of lactation days were considered and analyzed using a single-trait RRM, including the additive genetic, permanent environmental and residual random effects. In addition, contemporary group, linear and quadratic effects of the covariate age of the buffalo at birth, number of milkings and mean lactation curve of the population, modeled using a fourth-order orthogonal Legendre polynomials, were included as fixed effects. Covariance functions of additive genetic and permanent environmental regression coefficients were equally modeled by POL (3<sup>th</sup> to 6<sup>th</sup> order) or linear BPL (4 to 6 knots). Residual variances were modeled considering 9 classes uniformly spaced. The (co)variance components of parameters were estimated by Bayesian inference, using the GIBBS3F90 software. The convergence of Gibbs chains was verified by graphic inspection and using BOA package from the R program. The post-gibbs analyses were performed using the POSTGIBBS1F90 software. The selection of the models was based on deviance information criterion (DIC). The average TDMY was  $6.31 \pm 2.13$  kg. The sixth order POL and six-knot BSP (knots at 8, 29, 57, 120, 239 and 267 days), with highest degree of complexity (51 parameters) were considered as the most suitable to describe the TDMY variation. The genetic parameters presented similar tendencies for both models, excepted for genetic variances and heritability, which showed low estimates at first lactation weeks for model BSP6. It was possible realize that, despite the low correlation between the predicted breeding value for 270-day milk yield and lactation persistency, the best models classified the bulls similarly. The results indicate that linear B-spline function with six knots was the most appropriate RRM to genetic evaluation of TDMY of Murrah buffaloes,

given that parsimonious models with higher computational advantages to model variations of TDMY are most sought, and given the better control of heritability estimates of the edges of the lactation curve.

## 1- INTRODUÇÃO

A bubalinocultura tem apresentado altas taxas de crescimento no país, caracterizando-se como importante segmento da pecuária brasileira. A produção de leite destinada à fabricação de derivados lácteos tem se mostrado como uma das formas potencialmente rentáveis de exploração da espécie, conduzindo produtores a adotá-la em suas propriedades e a promoverem melhorias neste segmento para aumento da lucratividade (TONHATI, 2008).

As melhorias necessárias somente são perceptíveis a partir do conhecimento da situação produtiva dos plantéis, consequência do controle rígido dos dados produtivos dos animais. Assim, facilita e possibilita posterior inserção destas propriedades em programas de melhoramento genético desenvolvidas por instituições públicas e privadas.

A produção de leite se destaca como uma característica de grande importância para a atividade. As produções de leite no dia de controle são classificadas como características longitudinais, em que é perceptível uma estrutura entre os registros de produção de leite ao longo da lactação, dadas pelas covariâncias entre as mesmas e que devem ser consideradas na avaliação genética (MEYER, 2005a).

A referida estrutura de covariâncias pode ser modelada com base em modelos de regressão aleatória, comumente empregados para características longitudinais. Nesse grupo de modelos estão inseridos os polinômios ortogonais de Legendre, que possuem a vantagem de serem regressões ortogonais que proporcionam maior facilidade de convergência, decorrente da redução de correlação entre os coeficientes estimados. Os referidos polinômios, normalmente, necessitam de regressões de ordens elevadas para que se ajustem adequadamente à curva de lactação, que resulta na obtenção de estimativas de variâncias inadmissíveis nos extremos da lactação e correlações negativas entre semanas de controle mais longínquas (POOL et al., 2000; LÓPEZ-ROMERO et al., 2004).

Uma alternativa para diminuir a ordem dos polinômios é por meio da utilização de funções *B-splines* lineares, que consistem de polinômios segmentados independentes, conectados em pontos específicos denominados *nós* (DE BOOR, 1978). Os *B-splines* permitem a redução da multicolinearidade, são facilmente estimadas (linear nos

parâmetros), além de possuírem interpretação biológica simples e maior flexibilidade de adaptação à curva de lactação (WHITE et al., 1999; MISZTAL, 2006).

Uma das metodologias empregadas para estimação dos coeficientes das regressões segmentadas- e posteriormente os parâmetros de interesse - é a utilização de métodos bayesianos, com base em Cadeia de Markov Monte Carlo (MCMC). Esta metodologia, associada à Amostragem de Gibbs, resulta em menor custo computacional para estimação de componentes de (co)variância, além da maior flexibilidade quanto às inferências estatísticas que podem ser realizadas a partir dos resultados, como a obtenção de intervalos de credibilidade (VAN TASSEL et al., 1995).

Os *B-splines* lineares são funções que podem ser usadas em regressão aleatória para dados de produção de leite de bovinos. Vários estudos envolvendo a aplicação de funções *B-splines* para dados de bovinos leiteiros com o objetivo de determinar parâmetros genéticos e estimação dos pontos de nós ao longo da lactação são perceptíveis na literatura (BOHMANOVA et al., 2008; JAMROZIK et al., 2010; BIGNARDI et al., 2012). No entanto, em búfalos leiteiros, informações sobre a utilização de funções *B-splines* para estimativas de parâmetros genéticos, são incipientes.

## 2- OBJETIVOS

O objetivo proposto com este trabalho é estudar a variação da produção de leite no dia de controle de búfalas da raça Murrah por meio de modelos de regressão aleatória (MRA), utilizando polinômios ortogonais de Legendre (POL) e polinômios lineares segmentados do tipo B (*B-splines*).

Os objetivos específicos foram:

- a) Contrastar os MRA ajustados para diferentes números de segmentos lineares de polinômios do tipo B e diferentes ordens de polinômios ortogonais de Legendre e identificar o melhor modelo para a avaliação genética de bubalinos;
- b) Avaliar o comportamento das cadeias geradas via amostrador de Gibbs (convergência e número de amostras efetivas);
- c) Estimar componentes de (co)variância e parâmetros genéticos e respectivos intervalos de maior densidade posterior (HPD) ao longo da lactação;
- d) Predizer e relacionar valores genéticos para produção de leite aos 270 dias e persistência da lactação.

### 3- REVISÃO BIBLIOGRÁFICA

#### 3.1. Histórico e situação da bubalinocultura

Os búfalos (*Bubalus bubalis* Artiodactyla, Bovidae) são animais domésticos originários da Ásia, amplamente difundidos em vários continentes (RODRIGUES et. al., 2008). Segundo Santiago (2000), essa espécie foi introduzida no Brasil por volta de 1890, na Amazônia, por meio de refugiados oriundos da Guiana Francesa. O efetivo bubalino brasileiro cresceu 7,8% ao comparar-se 2011 a 2010, totalizando 1,3 milhões de cabeças, sendo que aproximadamente 56% destes são encontrados na região Norte do país. No entanto, deve-se destacar o maior crescimento do rebanho na região Sudeste, com aumento de 8,8% no mesmo período (IBGE, 2011).

A espécie compreende animais de pele escura com baixa densidade de glândulas sudoríparas e uma epiderme espessa, que dificultam sua adaptação a condições extremamente quentes e secas. No Brasil, são reconhecidas pela Associação Brasileira de Criadores de Búfalos quatro raças divididas em dois grupos principais: Búfalo-de-rio (*River Buffalo*), com  $2n=50$  cromossomos - entre as quais a raça Mediterrânea, Murrah e Jafarabadi; e Búfalo-do-pântano (*Swamp Buffalo*), os *Bubalus bubalis* var. kerebau, com  $2n=48$  cromossomos ou Carabao - sendo essa a única raça reconhecida para o grupo (A.B.C.B., 2012).

Inicialmente, os búfalos se destacaram pela sua docilidade, rusticidade e resistência às enfermidades e parasitoses. Sua versatilidade tanto para a produção de leite, carne e trabalho permite aumento desse efetivo populacional bubalino, fato este comprovado pelas estatísticas oficiais (FAO, 2000; TONHATI et al., 2007).

Apesar das melhorias significativas na cadeia produtiva, das diversas ações promocionais para profissionalização e aprimoramento das técnicas produtivas que tem proporcionado melhores índices de desempenhos dos rebanhos, a produtividade é aquém da almejada. O encarecimento dos insumos e a falta de iniciativa por parte dos produtores na adoção de tecnologias são os principais entraves encontrados, restringindo assim o potencial produtivo dos animais (BERNARDES, 2006).

Segundo Tonhatiet al. (2008), grande parte dos criadores tem optado pela exploração da produção de leite nessa espécie para posterior fabricação de subprodutos em decorrência desta se apresentar como uma das formas mais rentáveis. O leite de búfala, dadas suas características físico-químicas peculiares é matéria prima ideal para elaboração de diversos tipos de queijo, em que o mais tradicional é o *mozzarella*, originário da Itália no século XVI, produzida exclusivamente com leite de búfala (RODRIGUES et. al., 2008; A.B.C.B., 2012).

Assim, a fim de explorar racional e economicamente todo o potencial produtivo da espécie, são necessários conhecimentos básicos sobre os parâmetros genéticos da população, principalmente no que se refere às características produtivas como as produções de leite ao longo da curva de lactação.

### **3.2. Avaliação de dados longitudinais**

As produções de leite do dia de controle (PLDC) são classificadas como medidas repetidas ou dados longitudinais, e são obtidas periodicamente (geralmente semanal ou mensalmente) ao longo de uma escala contínua (dias de lactação). Sem perder de vista outras características de importância econômica, as PLDC merecem um tratamento estatístico distinto, em virtude da existência de um padrão de covariâncias bem estruturado, em que a correlação entre as mensurações apresenta-se inversamente proporcional ao intervalo entre elas. Assim, estudos baseados em metodologias que avaliam estes dados longitudinais têm sido frequentes (VAN DER WERF, 2001; DORNELES et al., 2009; COBUCI et al., 2011; MENEZES et al, 2011).

A forma de análise mais simples para dados longitudinais são os modelos de repetibilidade, que partem do pressuposto de que os registros pertencentes ao animal são expressões de uma mesma característica. Como consequência, assume-se que as correlações genéticas entre as medidas sejam iguais à unidade e que a (co) variância entre todos os pares de registro são de mesma magnitude, situação esta que não corresponde à realidade biológica das características pertencentes à classe dos dados longitudinais (VAN DER WERF, 2001; ALBUQUERQUE, 2004; BUXADERA, 2005).

Uma alternativa de avaliação consiste em considerar cada medida de produção como uma característica diferente e realizar uma análise conjunta em um modelo multivariável (DRUET et al., 2005; HAMMAMI et al., 2008; ASPILCUETA-BORQUIS, 2010). Em relação à forma de análise anterior, esse tipo de análise possui a vantagem de oferecer diferentes estimativas de correlação entre as medidas e entre lactações, sem assumir qualquer estrutura de covariância. Posteriormente, podem ser preditos valores genéticos e estimados parâmetros genéticos para cada característica, ou seja, para cada dia de lactação específico utilizando uma maior quantidade de informações disponíveis, proporcionando estimativas mais precisas e acuradas. No entanto, em casos em que se deseja avaliar inúmeras características, o que geralmente ocorre quando avaliadas as produções de leite no dia de controle, há necessidade da estimação de grande quantidade de parâmetros que dificultam o processamento das informações.

O uso de modelos de regressão aleatória surgiu como alternativa para superar o problema de superparametrização das análises multivariáveis. Para os modelos de regressão aleatória, são estimadas variâncias e covariâncias entre os coeficientes de regressão para os efeitos aleatórios, que dependem da função ajustada. Tais modelos são muito convenientes, pois permitem prever valores genéticos e estimar a produção e componentes de variância em qualquer idade dentro do intervalo analisado (HENDERSON JR., 1982; RESENDE et al., 2001; SILVA, 2011).

Nos modelos de regressão aleatória é necessário estimar curvas com uma parte fixa, uma aleatória ambiental e outra aleatória genética aditiva. A parte fixa é relativa ao efeito médio da população e a parte aleatória de ambiente evidencia os efeitos ambientais correlacionados entre as medições. As curvas relativas ao efeito aleatório genético-aditivo são desvios atribuídos ao valor genético dos animais e, de um modo geral, são desvios da produção dos animais em torno da curva fixa ajustada de modo a permitir a variação individual na forma da trajetória em um nível genético (HALENDER et al., 2010; SILVA, 2011).

Primeiramente para aplicação dos MRA popularizou-se a utilização de funções paramétricas, a exemplo de Wilink e Ali & Schaeffer (ALI e SCHAEFFER, 1987). Posteriormente, priorizou-se a utilização de funções não-paramétricas nos processos de estimação de funções de covariância, como os polinômios ortogonais de Legendre que, em

decorrência de serem regressões ortogonais, possuem a vantagem computacional de possuírem menor correlação entre os coeficientes estimados, facilitando a convergência (POOL et al., 2000; SCHAEFFER, 2004).

Polinômios de Legendre de altas ordens são flexíveis e se mostraram capazes de modelar adequadamente as médias e variâncias ao longo da escala contínua. Entretanto, estes polinômios impõem sua modelagem em regiões da curva que apresentam poucas observações, o que geralmente ocorre nos extremos da curva, proporcionando altas variâncias nesses pontos e correlações negativas entre os mais distantes dias do controle (KIRKPATRICK et al., 1990; STRABEL e MISZTAL, 1999; MEYER, 2005a). Uma alternativa para diminuir a ordem das funções empregadas são os polinômios segmentados (*splines*).

### **3.3. Polinômios segmentados**

Segundo Schumaker (2007) a terminologia “função *splines*” foi introduzida por Schoenberg, em 1946. São definidas como curvas constituídas por segmentos individuais de polinômios de grau reduzido que se unem em pontos específicos chamados nós. Em essência, se ajustam regressões separadas dentro das regiões entre os nós, e os nós ligam os ajustes de regressão segmentada. Logo, cada coeficiente de uma função *spline* atua somente em uma fração da trajetória estudada. O intuito da utilização de modelos *splines* é aumentar a flexibilidade de adaptação à curva de lactação, proporcionar uma redução da multicolinearidade, facilitar a estimação de parâmetros genéticos e a interpretação biológica (WHITE et al., 1999; MEYER, 2005b; MISZTAL, 2006).

Inúmeras ordens de polinômios podem ser assumidas para os segmentos que conjuntamente representarão o modelo *spline*. Segmentos lineares são as formas mais simples de aplicação da metodologia e são extensões da própria regressão linear, comumente estudadas em estatística. São empregadas como decorrência do próprio objetivo da análise simplificá-la o máximo possível. Em detrimento à necessidade de escolha das ordens dos polinômios, o posicionamento e quantidade de nós se tornam os principais fatores para estimação adequada das (co)variâncias entre os coeficientes dos polinômios. Maiores quantidades de nós são necessárias para aumentar a correlação entre

os mesmos e tornar a função estimada mais suave e acurada possível. No entanto, há um aumento da complexidade do modelo, em virtude do aumento do número de parâmetros a serem estimados (BOHMANOVA et al., 2008; JAMROZIK et al., 2010).

Para implementar os segmentos tipo *splines*, são necessárias funções bases locais, e são preferíveis as chamadas *B-splines* (B significa base) decorrente das suas melhores propriedades numéricas e computacionais quando comparadas a outras funções base (DE BOOR, 2001; SCHUMAKER, 2007). Os *B-splines* são construídos ajustando um polinômio entre os pares de nós. Os coeficientes dos *B-splines* são calculados como uma interpolação dos coeficientes entre 2 nós adjacentes e 0 para os outros nós. Tendo em vista o modelo freqüentemente aplicado em estudos de dados longitudinais em bovinos de leite, em que são considerados os efeitos genéticos-aditivos e de ambiente permanente, o modelo de regressão aleatória para *B-splines* lineares pode ser representado da seguinte forma:

$$y_{jk}(t) = \mathbf{f}_j + \sum_{i=1}^n \varphi_i(t)a_{ij} + \sum_{i=1}^n \varphi_i(t)p_{ij} + e_{jk}$$

em que  $y$  é a observação  $k$  do animal  $j$  registrada no tempo  $t$ ;  $\varphi_i(t)$  é a covariável  $i$  no tempo  $t$ ;  $a_{ij}$  e  $p_{ij}$  correspondem aos coeficiente de regressão aleatório genético  $i$  e de ambiente permanente do animal  $j$ , respectivamente;  $\mathbf{f}_j$  corresponde ao conjunto de efeitos fixos inerentes ao animal  $j$ ; e  $e_{jk}$  é o resíduo para o registro  $k$  do animal  $j$ . Assumindo  $n$  nós nos pontos  $T_i$  ( $i = 1, \dots, n$ ), e que  $T_i \leq t < T_{i+1}$ , tem-se que as covariáveis podem ser obtidas, respectivamente nos nós  $i$  e  $i + 1$ :

$$\begin{aligned} \varphi_i(t) &= \frac{T_{i+1} - t}{T_{i+1} - T_i} \quad e \\ \varphi_{i+1}(t) &= 1 - \frac{T_{i+1} - t}{T_{i+1} - T_i} \quad . \end{aligned}$$

Para outros valores de  $i$ , tem-se:

$$\varphi_{outros}(t) = 0 \quad .$$

Vários estudos utilizando funções *B-splines* foram realizados em bovinos leiteiros (BOHMANOVA et al., 2008; JAMROZIK et al., 2010; BIGNARDI et al., 2012). No

entanto, em búfalos leiteiros, ainda não se encontram reportes sobre a utilização de funções *B-splines*, somente o uso de polinômios ortogonais de Legendre (BREDA et. al., 2010; SESANA et. al., 2010, ASPILCUETA-BORQUIS et. al., 2012).

### 3.4. Inferência bayesiana

A teoria bayesiana se fundamenta em um teorema proposto por Thomas Bayes, em 1761, o qual propõe caracterizar a aprendizagem como experiência. Sua filosofia emprega que o desconhecido – no caso em questão, o parâmetro a ser estimado  $\theta$  – é incerto e essa incerteza deve ser quantificada em termos de probabilidade (MURTEIRA, 1990), de forma que se pode apresentar:

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

em que  $p(\theta)$  é conhecida como distribuição de probabilidade *a priori* de  $\theta$ ;  $p(y|\theta)$  corresponde à distribuição condicional de  $y$ , dado o parâmetro  $\theta$  (proporcional à função de verossimilhança); e  $p(\theta|y)$  é a distribuição de probabilidade *a posteriori* de  $\theta$ .

Verifica-se dessa forma que a informação da distribuição *a priori* e a distribuição proporcional à função de verossimilhança são combinadas culminando na distribuição *a posteriori* de  $\theta$ . Esta combinação permite ampliar o conhecimento do estatístico sobre o comportamento da variável ao incorporar toda a informação disponível, funcionando assim como um mecanismo de atualização da opinião do avaliador sobre o parâmetro  $\theta$ . Como consequência, a maior disponibilidade de dados faz com que o efeito da “função de verossimilhança” sobreponha o efeito da distribuição *a priori*. As inferências estatísticas sobre os parâmetros são feitas com base na manipulação da distribuição *a posteriori*, possibilitando a obtenção de estimativas pontuais (a exemplo da moda ou média) e “intervalos de credibilidade” para os parâmetros, a exemplo do intervalo de maior densidade posterior (GIANOLA e FERNANDO, 1986; BEAUMONT e RANNALA, 2004).

As técnicas Bayesianas foram abandonadas durante muitos anos por requererem a resolução de integrais múltiplas extremamente complexas. A inferência exata somente seria possível se estas integrais pudessem ser calculadas analiticamente por meio de

métodos não iterativos, o que é impraticável. Como alternativa, utiliza-se métodos iterativos que possibilitem a solução de vários problemas decorrentes da impossibilidade de resolução dessas integrais, dentre as quais se destaca a técnica de Monte Carlo baseada em cadeias de Markov (**MCMC** – Markov Chain Monte Carlo) (BLASCO, 2001; EHLERS, 2007).

A técnica de Monte Carlo via Cadeias de Markov é um método de simulação para obtenção aproximada da solução das integrais comentadas anteriormente. Este método se dá de forma iterativa, ou seja, repetidamente são estimados valores para o parâmetro e são obtidas amostras aleatórias dos mesmos, que tendem a se aproximar da distribuição estacionária; a própria distribuição de probabilidade *a posteriori* (SORENSEN; GIANOLA, 2002).

GEMAN e GEMAN (1984) definem a amostragem de Gibbs, um caso particular do algoritmo de Metropolis-Hastings do método de MCMC, como um dos métodos de amostragem de  $\theta$ . Cada iteração do amostrador de Gibbs consiste em simular os valores das distribuições condicionais completas *a posteriori* de cada subvetor de  $\theta$  e, à medida que o número de iterações aumenta, a cadeia se aproxima da sua condição de equilíbrio.

VAN TASSEL et al. (1995) resumem as seguintes vantagens do uso do método da Amostragem de Gibbs em relação aos métodos usuais que empregam teoria BLUP: a) não requer soluções para as equações de modelo misto; b) permite a análise do conjunto de dados maiores do que quando se usa REML com técnicas de matrizes esparsas; c) propicia estimativas diretas e acuradas dos componentes de (co)variância, valores genéticos e intervalos de credibilidade para essas estimativas; e d) pode ser usado sem problemas em microcomputadores e estações de trabalho.

Uma grande vantagem desses métodos de simulação é que devido sua propriedade de obtenção indireta de variáveis aleatórias, a partir das amostras obtidas, podem ser calculadas características de distribuições marginais posteriores relativas aos parâmetros (à exemplo da média, moda, mediana e variância), tornando supérflua a necessidade de sua obtenção. Dessa forma, é possível visualizar problemas que não são perceptíveis utilizando metodologia de Máxima Verossimilhança Restrita (**REML**) (CASELLA e GEORGE, 1992).

## 4- METODOLOGIA

### 4.1. Dados

O banco de dados utilizado foi proveniente de 12 rebanhos localizados no estado de São Paulo, Brasil, incluídos no Programa de Controle Leiteiro de Bubalinos mantido pelo Departamento de Zootecnia da Faculdade de Ciências Agrárias e Veterinárias (FCAV) da Universidade Estadual Paulista (UNESP) “Julio de Mesquita Filho” – Câmpus de Jaboticabal, São Paulo, com partos obtidos no período de 1988 a 2005. Os rebanhos são formados por animais da raça Murrah (*Bubalus bubalis*). Os registros de produção de leite variaram de acordo com o sistema de controle leiteiro adotado pela propriedade (semanal ou mensal).

A organização e consistência do arquivo de dados para as análises foram realizadas utilizando o pacote estatístico SAS 9.1 (2003). Para ajuste dos modelos de regressão aleatória, as produções de leite foram consideradas em classes semanais de lactação, sendo utilizados registros obtidos a partir do 5º ao 270º dia de lactação. As semanas foram contabilizadas a partir do 5º dia após o parto, ou seja, a primeira semana de lactação corresponde ao intervalo do 5º ao 11º dia de produção de leite da búfala, e assim por diante.

Somente foram consideradas informações de animais que atenderam requisitos pré-estabelecidos, dentre os quais: a) lactações maiores que 90 dias; b) primeiro registro até o 45º dia após o parto; c) intervalo de controles de, no máximo, 45 dias; d) idades ao primeiro parto entre 24 e 60 meses; e e) a búfala deveria possuir, ao menos, 5 (cinco) registros de produção de leite no dia de controle (PLDC).

Os grupos de contemporâneos (GC) foram formados por animais pertencentes ao mesmo rebanho e ao mesmo ano e mês de controle. Desconsiderou-se GC com menos de 5 (cinco) animais, assim como registros de produção de leite que extrapolaram 3,5 desvios-padrão acima ou abaixo da média de PLDC dentro de grupo de contemporâneos e dentro da classe semanal. Ao final da etapa de consistência, o arquivo utilizado nas análises possuía 15161 registros de PLDC de 1158 primeiras lactações e um arquivo de pedigree contendo 1852 indivíduos.

## 4.2. Modelos

A produção do leite no dia de controle foi analisada por meio de um modelo animal de regressão aleatória. Os efeitos de grupo de contemporâneo, o número de ordenhas por dia (1 ou 2 ordenhas diárias), a covariável idade da búfala ao parto (efeito linear e quadrático) e as médias dos coeficientes da curva de lactação que representa a população - modelada por meio de um polinômio ortogonal de Legendre de quarta ordem - foram considerados como efeitos fixos. Para todos os modelos foram incluídos os efeitos aleatórios genético-aditivo, de ambiente permanente de animal e resíduo. A estrutura de variâncias residuais foi considerada heterogênea, contendo 9 (nove) classes (1-4, 5-8, 9-12, 13-16, 17-20, 21-24, 25-28, 29-32, 33-38 semana de lactação).

Para modelagem das funções de covariâncias para os efeitos genético-aditivos e de ambiente permanente foram utilizados polinômios ortogonais de Legendre (**POL**) ou “*B-splines*” lineares (**BSP**). A nomenclatura utilizada para os modelos segue a seguinte notação: POLX e BSPY, de forma que X representa a ordem do polinômio empregando (3 a 6); e Y representa o número de nós (4 a 6) utilizados na definição dos *B-splines*. Neste trabalho, os efeitos genético-aditivos e de ambiente permanente foram modelados empregando a mesma ordem de POL ou número de nós, no caso dos BSP.

No que se referem aos modelos empregando BSP, os  $m$  nós foram escolhidos para dividir as semanas de lactação em  $m-1$  intervalos e os nós externos foram localizados no início e no encerramento da lactação para todos os modelos (1ª e 38ª semana). O primeiro nó interno foi posicionado próximo à região do pico de lactação (57º dia – 8ª semana), sendo considerados mais dois nós posteriores na 17ª (120 dias) e 34ª (239º dia) semana, conforme avaliação da estrutura dos dados e da curva de lactação. Para o último modelo, um nó anterior ao pico de lactação foi incluído (4ª semana), de forma a permitir maior maleabilidade ao modelo.

As estimativas dos componentes de (co)variâncias foram obtidas via inferência Bayesiana, utilizando-se o programa computacional GIBBS3F90 (MISZTAL, 2012), no qual encontra-se implementado o método de amostragem de Gibbs.

O modelo animal de regressão aleatória utilizado pode ser representado matricialmente como:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{ap} + \mathbf{e}$$

em que  $\mathbf{y}$  = vetor das N observações, medidas em  $N_d$  animais;  $\mathbf{b}$  = vetor de efeitos “fixos” que inclui as soluções para o grupo contemporâneo (612 níveis), para a covariável idade da vaca ao parto (regressão linear e quadrática), número de ordenhas diárias (2 níveis) e as médias dos coeficientes relativas a população;  $\mathbf{a}$  e  $\mathbf{ap}$  = vetores de soluções para os coeficientes de regressão aleatória genético-aditivo e de ambiente permanente, respectivamente;  $\mathbf{e}$  = vetor dos resíduos;  $\mathbf{X}$  = matriz de incidência que associa o vetor de efeitos fixos ( $\mathbf{b}$ ) ao vetor de observações ( $\mathbf{y}$ );  $\mathbf{Z}$  e  $\mathbf{W}$  = matrizes de covariáveis referentes às idades em cada produção modeladas pela função (polinômio ortogonal de Legendre ou *B-splines* lineares) que descreve a trajetória, associadas aos coeficientes de regressão aleatória dos efeitos aleatório genético-aditivo ( $\mathbf{a}$ ) e de ambiente permanente ( $\mathbf{ap}$ ), respectivamente.

As pressuposições acerca das distribuições *a priori* para os parâmetros foram consideradas como a seguir:

- $\mathbf{a} | \Lambda_g \sim NMV(0, G)$ , com  $G = \Lambda_g \otimes A$ , em que  $A$  é a matriz dos numeradores dos coeficientes de parentesco de Wright e  $\Lambda_g$  é a matriz contendo as covariâncias entre os coeficientes de regressão aleatória genético-aditivos;
- $\mathbf{p} | \Lambda_{ap} \sim NMV(0, P)$  com  $P = \Lambda_{ap} \otimes I$ , em que  $I$  é uma matriz identidade de dimensão  $N_d$  e  $\Lambda_p$  é a matriz contendo as covariâncias entre os coeficientes de regressão aleatória de ambiente permanente;
- $\Lambda_g | v_g, S_g^2 \sim W^{-1}(v_g, v_g S_g^2)$ ,  $\Lambda_{ap} | v_{ap}, S_{ap}^2 \sim W^{-1}(v_{ap}, v_{ap} S_{ap}^2)$  em que  $v_g S_g^2$  e  $v_{ap} S_{ap}^2$  representam o grau de confiança e os valores das distribuições *a priori* para as covariâncias dos coeficientes de regressão genético-aditivos e de ambiente permanente, respectivamente;
- $\sigma_{ei}^2 | v_e, S_e^2 \sim \chi^{-2}(v_e, v_e S_e^2)$ , em que  $v_e S_e^2$  representam o grau de confiança e os valores da distribuição *a priori* para a variância residual da classe  $i$ , e;

- $p(b) \propto \text{constante}$ , representando a falta de conhecimento da distribuição *a priori* sobre os efeitos.

$NMV$ ,  $W^{-1}e\chi^{-2}$  referem-se às distribuições normal multivariada, *Wishart* invertida e chi-quadrado invertida, respectivamente.

As inferências sobre os parâmetros de interesse foram realizadas a partir das correspondentes amostras das distribuições marginais posteriores. Para cada modelo avaliado foi gerada uma cadeia com 650.000 iterações, com um descarte inicial (*burn-in*) de 100.000 amostras - sendo determinado um intervalo de coleta (*thin*) de 100 amostras, totalizando assim 5500 amostras para a realização das inferências. Os componentes de (co)variância e parâmetros genéticos para cada PLDC foram obtidas vetor a vetor, sendo geradas distribuições posteriores para todos os parâmetros em estudo, de forma que:

$$V_{ni} = \Psi \Lambda_{ni} \Psi'$$

em que  $V$  corresponde a matriz de (co)variâncias entre as produções de leite no dia de controle, de ordem  $m$ ;  $\Psi$  é a matriz de covariáveis da função base (polinômios ortogonais de Legendre ou *B-splines*), cuja dimensão é  $m \times p$ ;  $\Lambda$  é a matriz de ordem  $p$  de estimativas de (co)variâncias entre os coeficientes de regressão aleatória do efeito aleatório  $n$  da iteração  $i$  da cadeia de Gibbs;  $m$  é o número de classes de PLDC (9 classes); e  $p$  é a ordem do polinômio ou o número de segmentos.

A verificação da convergência das cadeias geradas para os parâmetros (coeficientes de regressão e herdabilidade) baseou-se em análise gráfica e métodos de Heidelberger & Welch (1983) e Geweke (1992), disponíveis no pacote *Bayesian Output Analysis - BOA* (SMITH, 2007) - disponível na plataforma estatística R. Posteriormente à obtenção dos coeficientes da função de covariância para cada componente foram calculadas as variâncias e seus respectivos intervalos de alta densidade posterior (HPD95%), admitindo-se todas as amostras possíveis. Na etapa pós-gibbs foi utilizado o programa POSTGIBBSF90 (MISZTAL, 2012).

### 4.3. Comparação de modelos

Os modelos de regressão aleatória testados foram classificados com base na adoção do critério de informação da *deviance* bayesiana. Para efeito comparativo, foram estimados os valores genéticos para as características produções de leite acumulada aos 270 dias e persistência da lactação para os modelos selecionados e associação entre os mesmos.

O **Critério de Informação da Deviance (DIC)**, apresentado por Spiegelhalter et al. (2002), é representado da seguinte maneira:

$$DIC = \bar{D}(\theta) + p_D = 2\bar{D}(\theta) - D(\bar{\theta})$$

em que  $\bar{D}(\theta) = E_{\theta|y}[D(\theta)]$  (esperança da distribuição *a posteriori* da *deviance* bayesiana) e  $D(\bar{\theta}) = -2\log p(y|\theta)$  representam a qualidade de ajuste do modelo. O número efetivo de parâmetros representa a penalização para o incremento da complexidade do modelo:  $p_D = \bar{D}(\theta) - D(\bar{\theta})$ , onde  $\theta$  é o vetor de parâmetros do modelo e  $D(\bar{\theta})$  é a *deviance* bayesiana avaliada com base nas médias da distribuição *a posteriori* dos parâmetros.

Como consequência, foram preditos os valores genéticos para todos os animais (**VG** ou **EBV**, do inglês *Expected Breeding Value*), vetor a vetor, de forma a se obter distribuições posteriores destes parâmetros no modelo. Todavia, foram observados apenas os valores dos touros que possuíam menor, médio e maior número de filhas, para produção de leite acumulada aos 270 dias (**PL270**) e persistência da lactação (**PERSL**). Os modelos foram comparados utilizando correlações de *Pearson* e visualização do comportamento dos valores genéticos.

Os valores genéticos preditos para PL270 foram calculados por meio da soma de todos os valores genéticos preditos para as PLDC, como indicado por Tonhati et al. (2004). O cálculo para obtenção da persistência de lactação, por sua vez, baseou-se em uma modificação da função desenvolvida por Jamrozik et. al. (1997) e testada por Chaves (2009) em búfalas, que constataram baixa correlação desta com a produção de leite acumulada aos 305 dias e a escolha de menores valores para tal característica indicam animais com maiores persistências. Sendo assim, tem-se que:

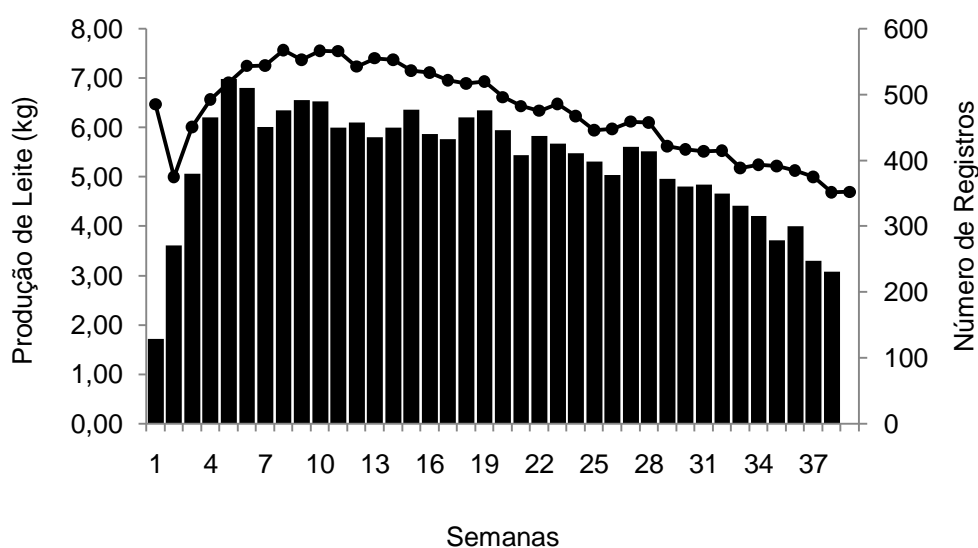
$$PERSL = \sum_{t=58}^{270} (VG_t - VG_{57})$$

em que  $VG_t$  e  $VG_{57}$  correspondem aos valores genéticos preditos no dia  $t$  e aos 57 dias em produção.

## 5- RESULTADOS E DISCUSSÃO

### 5.1. Estrutura dos dados

Na Figura 1 são apresentados o número de observações e média para a característica PLDC em função da semana de lactação. A PLDC apresentou média de  $6,31 \pm 2,13$  Kg. De forma geral, observou-se aumento da produção de leite até atingir o pico de lactação, por volta do 60º dia (8ª semana) e posterior decréscimo com o avanço da lactação.



**Figura 1** – Número de registros (barra) e média (•) das produções de leite observadas em função da semana de controle das primeiras lactações de búfalas da raça Murrah.

Os valores apresentados se encontram dentro dos padrões aceitáveis para bubalinos de aptidão leiteira, com exceção da 1ª semana, em que há um acréscimo atípico da média de PLDC. Silva et al. (2010) e Hurtado-Lugo et al. (2006), ao utilizarem dados de

produção de leite de bubalinos criados no nordeste do Brasil e na Colômbia, respectivamente, reportaram produções de leite inferiores às obtidas no presente estudo.

## 5.2. Convergência e descrição da cadeia de Gibbs

Análises preliminares demonstraram que, ao utilizar o modelo tendo como função base *B-spline*, a convergência foi atingida em menor quantidade de ciclos quando comparada aos POL (resultados não mostrados). No entanto, constatou-se que ao serem eliminados os primeiros 100.000 ciclos todos os modelos apresentaram convergência, sendo então esse *burn-in* adotado para todas as análises. O fato foi observado também por Misztal (2006) e Jamrozik (2010), que atribuíram o ocorrido a muitos fatores, dentre os principais: às melhores propriedades numéricas dos *B-splines*, haja vista que os coeficientes obtidos afetam apenas uma fração da trajetória; e o fato de que apenas as covariáveis associadas a dois nós adjacentes são diferentes de zero, tornando as matrizes mais esparsas, assim facilitando manipulação das mesmas.

A convergência da cadeia e, conseqüentemente, a estimação dos parâmetros são afetadas por diversos fatores, dentre as quais são possíveis citar a estrutura do banco de dados e a complexidade dos modelos. Adicionalmente são encontradas dificuldades para que a cadeia atinja o ponto de equilíbrio decorrente da influência das correlações entre os parâmetros a serem estimados e pela autocorrelação entre amostras, que por sua vez é inversamente proporcional ao número de amostras independentes. O número de amostras independentes (NAI) para as estimativas dos componentes de (co)variância alteraram entre modelos e efeitos testados (Tabela 1) (JAMROZIK, 2004; BOLIGON et al., 2011; EHLERS, 2007). É possível observar decréscimo dos NAI para o modelo POL4, possivelmente decorrente da insuficiente complexidade do presente modelo para ajuste dos dados.

**Tabela 1** – Média (Méd), mediana (Medi), mínimo (Mín) e máximo (Máx) do número de amostras independentes para os coeficientes dos componentes aleatórios genético-aditivos (G), ambiente permanente (PE) e residual para os modelos testados

Mod	G				PE				RESIDUAL			
	Méd	Medi	Mín	Máx	Méd	Medi	Mín	Máx	Méd	Medi	Mín	Máx
POL3	340	69	22	1464	489	93	33	1918	5428	5500	5165	5500
POL4	52	23	6	345	84	26	4	532	4628	5157	172	5500
POL5	941	1017	61	2858	964	1004	54	1695	5268	5406	4787	5500
POL6	557	490	148	1070	651	629	226	1412	5250	5352	4630	5500
BSP4	1654	1628	49	3312	2058	2092	53	4255	4892	5000	4575	5000
BSP5	950	890	446	1668	1290	1225	424	2468	4690	4720	4376	5000
BSP6	406	382	18	1055	654	547	71	1599	4945	5358	3271	5500

Em média, o número de amostras independentes, também denominado tamanho de amostra efetiva, correspondeu a 13,4% do número total de amostras (5500) para os efeitos genético-aditivo, 17,1% para os efeitos de ambiente permanente e 90,8% para o efeito residual. O comportamento de maior número de amostras independentes para o componente residual, seguido pelos de ambiente permanente e genético-aditivos, corroboram com os resultados obtidos por Jamrozik (2004), Pereira et al. (2013) e Strabel e Jamrozik (2006); e são, de uma forma geral, consequência das pressuposições assumidas para os efeitos: independência entre classes de resíduos, covariâncias nulas entre medidas obtidas em animais diferentes (para o efeito de ambiente permanente) e consideração do relacionamento genético entre os animais, para o efeito genético-aditivo.

## 5.2. Comparação de modelos

No que se refere à comparação dos modelos (Tabela 2), os que apresentaram maior complexidade, conforme necessidade de estimação de maior número de parâmetros, tenderam a apresentar melhores valores de *deviance* bayesiana (parâmetro para definição do grau de ajuste do modelo) conforme esperado. Ao se considerar o mesmo número de parâmetros entre modelos empregando *B-splines* e polinômios ortogonais de Legendre, verificou-se que na determinação dos valores de DIC o número de parâmetros efetivos penalizou os POL quando em modelos de maiores ordens (POL5 e POL6), em comparação com modelos BSP.

**Tabela 2** – Posição dos nós (PN), número de parâmetros (NP), número de parâmetros efetivos (NPE), média posterior da *deviance* bayesiana ( $\bar{D}(\theta)$ ), *deviance* bayesiana avaliada com base nas médias da distribuição *a posteriori* dos parâmetros ( $D(\bar{\theta})$ ), critério de informação da *deviance* (DIC) para os modelos utilizando polinômios ortogonais de Legendre (POL) e *B-splines* (BSP)

MOD	PN (dias)	NP	NPE	$\bar{D}(\theta)$	$D(\bar{\theta})$	DIC	Segundos/ Iteração
POL3	-	21	2746,44	44262,73	41516,29	47009,17	0,0827
POL4	-	29	2952,00	43895,80	40943,80	46847,80	0,1005
POL5	-	39	3406,98	43171,10	39764,12	46578,08	0,1193
<b>POL6</b>	-	<b>51</b>	<b>3360,04</b>	<b>42833,70</b>	<b>39473,66</b>	<b>46193,74</b>	<b>0,1370</b>
BSP4	8, 57, 120, 267	29	2978,91	43687,96	40709,05	46666,86	0,1008
BSP5	8, 57, 120, 239, 267	39	3281,24	43450,31	40169,07	46731,54	0,1169
<b>BSP6</b>	<b>8, 29, 57, 120,239, 267</b>	<b>51</b>	<b>3227,93</b>	<b>43185,07</b>	<b>39957,15</b>	<b>46413,00</b>	<b>0,1318</b>

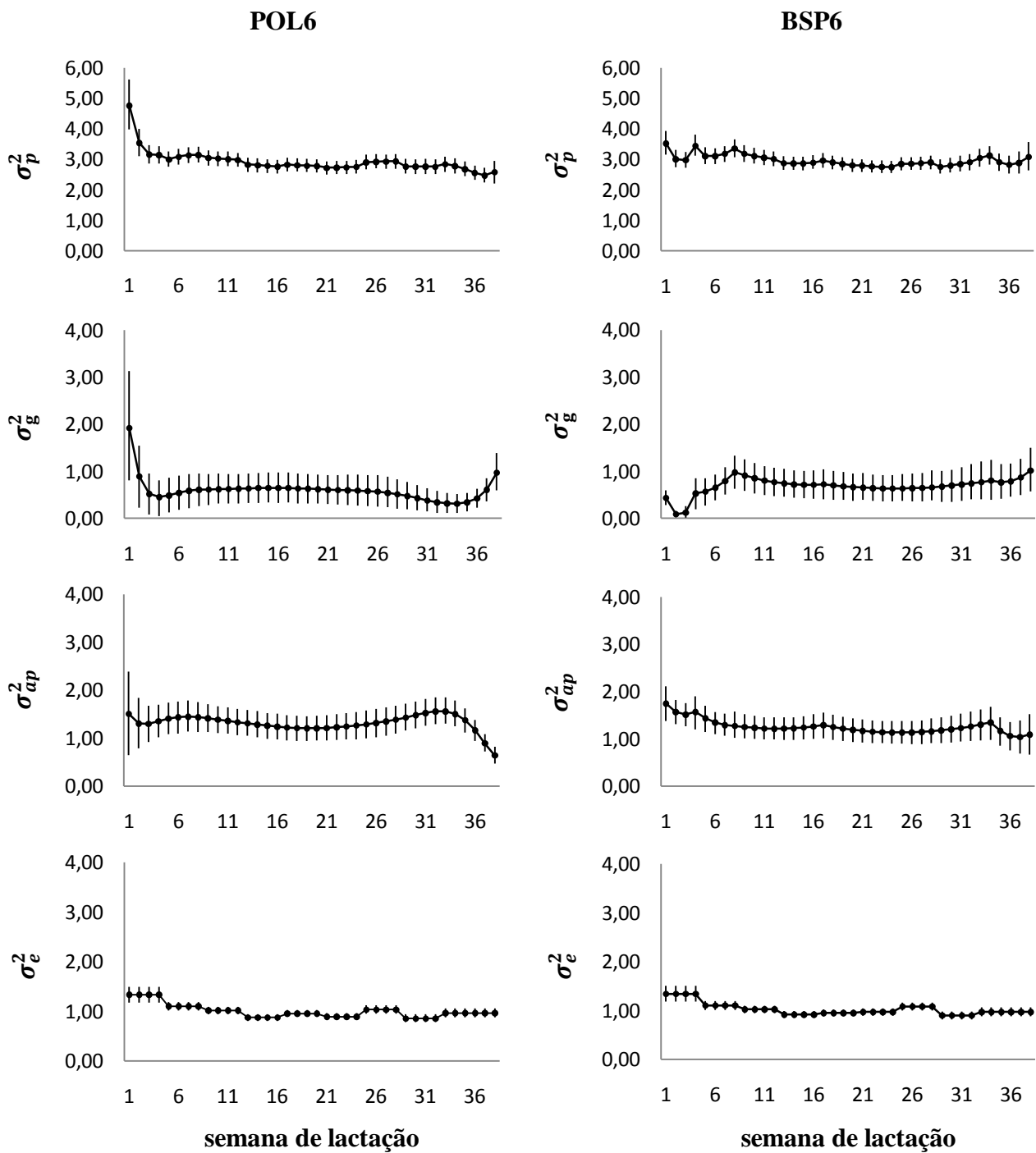
Com base nas estimativas de DIC, os melhores modelos foram aqueles que apresentaram maior número de parâmetros (POL6 e BSP6), mesmo considerando que quanto maior este fator o critério tende a penalizá-los. E, dentre eles, o modelo utilizando polinômios ortogonais de Legendre (POL6) se destacou. Acredita-se que a partição de classes de controle torna necessária a utilização de modelos mais complexos haja vista que, quando comparado a estudos mensais com a espécie (ASPILCUETA-BORQUIS et al., 2012), modelos com ajustes mais adequados apresentavam apenas 20 parâmetros. Outra possível explicação para a seleção de BSP6 e POL6 pelo critério adotado é o desbalanceamento do número de registros, que impede um ajuste adequado de modelos menos parametrizados (MEYER, 2005b).

Conforme esperado, os modelos mais complexos apresentaram maior tempo gasto por iteração. Apesar da diminuta extensão, a proporcionalidade de aumento do tempo por iteração conforme aumento do número de parâmetros foi menor quando empregado os modelos BSP. Esse fato remete a um menor tempo total gasto durante as análises para os modelos BSP - quando utilizado modelos mais complexos. Resultados similares foram encontrados por Bohmanova et al. (2008), que atribuiu o ocorrido devido às maiores correlações entre os nós quando comparada às correlações existentes entre os coeficientes dos POL. Segundo Misztal (2006), ao realizar análises que empregam *B-splines*, a maior

proporção de uso de matrizes mais esparsas utilizadas nas equações de modelos mistos proporcionam inúmeras vantagens, dentre as quais pode-se citar o menor tempo apresentado necessário por iteração.

### **5.3. Descrição das estimativas dos parâmetros**

As variâncias fenotípicas para o modelo POL6 (Figura 2) foram maiores no início da lactação quando comparadas às estimadas pelo BSP6, permanecendo estáveis a partir da quarta semana. Tal comportamento pode ser explicado pelo fato de que nos extremos da curva os efeitos ambientais apresentam uma maior influência na produção de leite da búfala. Dentre os quais é possível citar o estresse decorrente do parto, a capacidade inicial das células secretoras da glândula mamária e tratamentos diferenciados entre animais dentro e entre rebanhos no intervalo de partos (JAMROZIK et al., 2001; BOHMANOVA et al., 2008; PENCHEV et al., 2011; BIGNARDI et al., 2012). Diferenças entre variâncias fenotípicas no início da lactação também foram constatadas Bignardi et al. (2011) e Pereira et al. (2013).



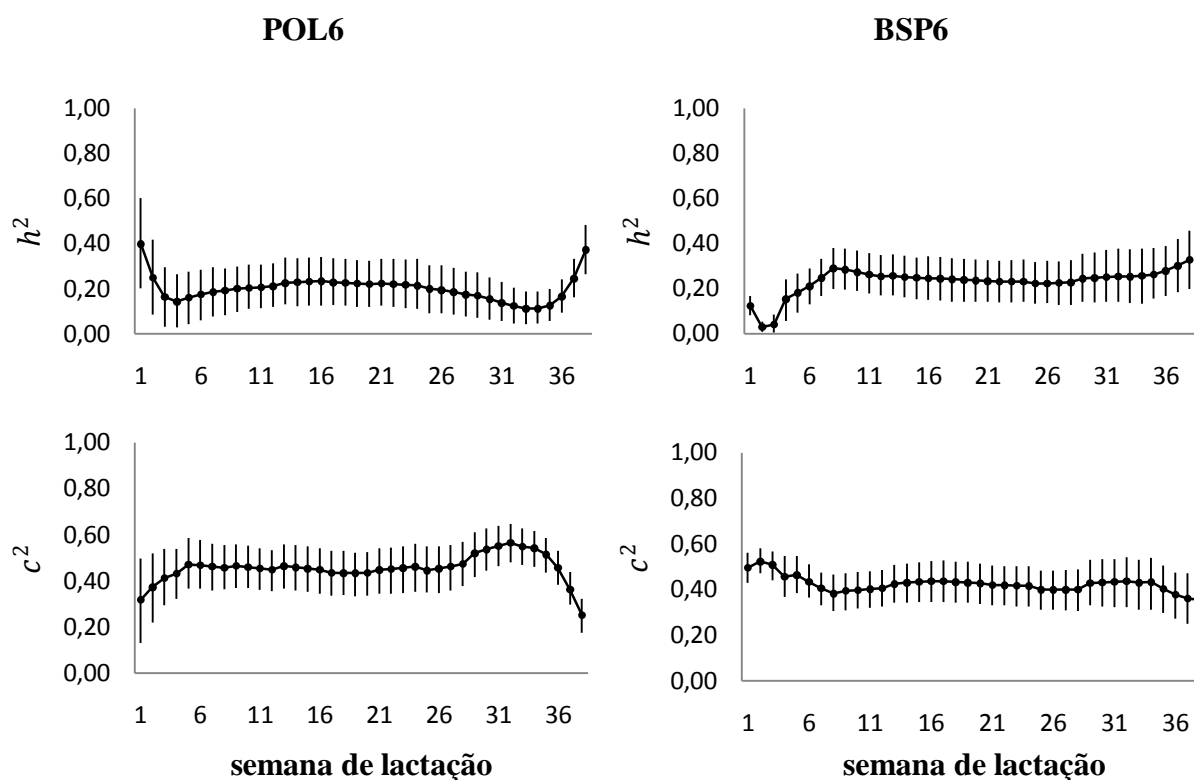
**Figura 2** – Médias (•) e 95% do intervalo de maior densidade posterior (barra) das distribuições posteriores das estimativas de variância fenotípica ( $\sigma_p^2$ ), genética aditiva ( $\sigma_g^2$ ), de ambiente permanente ( $\sigma_{ap}^2$ ) e residual ( $\sigma_e^2$ ) para os modelos POL6 e BSP6 em função da semana de lactação.

Com base nos intervalos, é possível notar que as estimativas de variância genética aditiva e de ambiente permanente nos pontos iniciais da curva tenderam a flutuar consideravelmente para os POL6. É perceptível que ao utilizar polinômios ortogonais de Legendre a variância genética aditiva foi superior nas primeiras semanas de lactação, quando comparadas aos BSP6. Ao empregar POL, frequentemente, são obtidas estimativas improváveis e duvidosas nos extremos das curvas representativas dos componentes aleatórios - em decorrência da alta ordem destes polinômios atribuírem grande ênfase nas observações destas regiões e, conseqüentemente, ampliam a susceptibilidade destas estimativas à modelagem imposta por estes polinômios (MEYER, 2005b; MISZTAL, 2006).

Para uma modelagem adequada empregando polinômios segmentados, aconselha-se inserção de nós adicionais em regiões da curva onde se verificam modificações abruptas nas estimativas de variâncias e quando estas se mostrarem demasiadamente irregulares ao longo da curva. Quando essas alternativas se mostrarem insuficientes, há ainda possibilidade de uso de transformações das covariáveis para tornar mais constante/linear essa variância entre nós (MEYER, 2005b; JAMROZIK et al., 2010; KIRKPATRICK et al., 1990).

Ao término da lactação há um decréscimo nas estimativas de variância de ambiente permanente, com diferenças estatísticas entre os modelos (verificados pelos intervalos). As variâncias residuais, no entanto, são bem similares entre as classes; a exemplo da região intermediária da curva projetada para o modelo BSP6, indicativo da possibilidade de junção entre as mesmas quando avaliados controles semanais, permitindo assim uma redução no número de classes de variância residual.

De uma forma geral, as estimativas das herdabilidades apresentaram uma tendência de crescimento e a porção da variação fenotípica atribuída ao ambiente permanente uma tendência decrescente ao empregar *B-splines* como função base. As médias das distribuições posteriores das herdabilidades variaram entre 0,11 e 0,40 (POL6) e 0,03 e 0,33 (BSP6), e para a porção de ambiente permanente variaram entre 0,25 a 0,57 e 0,36 a 0,52, para o modelo POL6 e BSP6, respectivamente (Figura 3).



**Figura 3** – Médias (•) e 95% do intervalo de maior densidade posterior (barra) das distribuições posteriores das estimativas de herdabilidade ( $h^2$ ) e porção da variação fenotípica atribuída ao ambiente permanente ( $c^2$ ) para os modelos POL6 e BSP6 em função da semana de lactação.

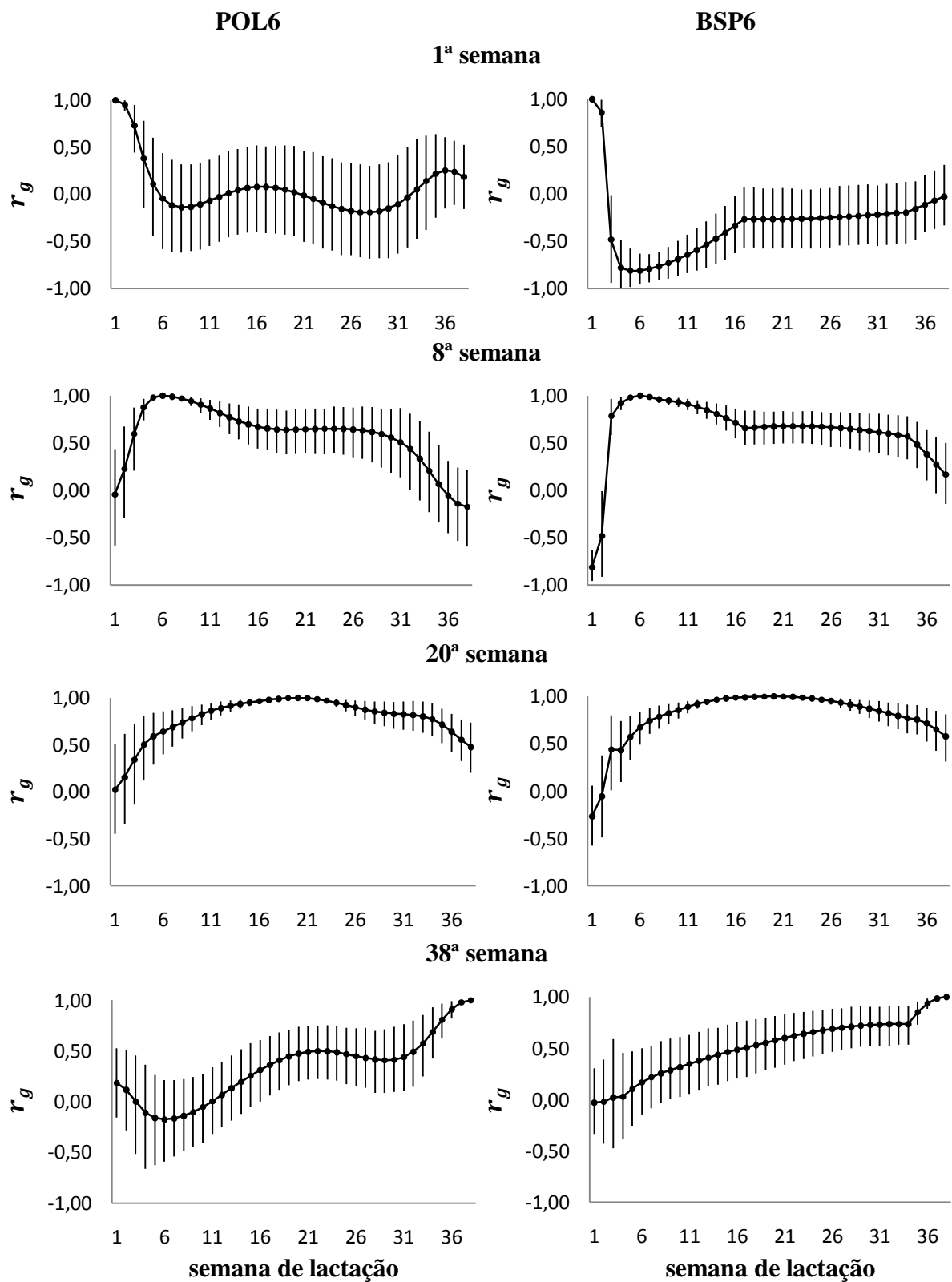
À exceção das estimativas de herdabilidade nas três primeiras semanas para o modelo BSP6, os valores foram próximos aos encontrados por Aspilcueta-Borquis et al. (2012), com estimativas de herdabilidade variando de 0,16 a 0,29 para controles mensais; e aos reportados por Breda et al. (2010), com valores entre 0,19 a 0,44 em controles semanais em búfalas- criadas no estado de São Paulo. Araújo et al. (2007), utilizando-se de registros de produções de leite no dia de controle de búfalos na região norte, obteve estimativas de herdabilidade oscilando entre 0,08 a 0,40, amplitude próxima aos valores encontrados neste trabalho.

Ambos os modelos apresentaram entre as semanas 4 e 28 estimativas de herdabilidades próximas: 0,14-0,17 (POL6) e 0,15-0,22 (BSP6). No entanto, nas três primeiras semanas verificaram-se maiores estimativas para o modelo POL6 (0,16-0,39) em comparação ao modelo BSP6 (0,03-0,12). A partir da 29ª semana, a maioria das

estimativas obtidas para BSP6 foram superiores em relação ao modelo utilizando polinômios de Legendre, mantendo assim uma maior constância destas estimativas ao longo das semanas de lactação para o modelo apresentado.

Bignardi et al. (2011) e Pereira et al. (2013), trabalhando com produções de leite em animais da raça Holandesa e Gir, respectivamente, descreveram diferenças nas estimativas de herdabilidade entre POL e BSP apenas nas primeiras semanas. Essas herdabilidades sugerem que os *B-splines* são menos susceptíveis às estimativas implausíveis, geralmente superestimadas para os efeitos nos extremos da curva, frequentemente observadas em análises utilizando POL. Meyer (2005b) atribui o fato à influência local das observações individuais dos modelos e à influência global das observações quando empregada POL.

Verificou-se que as estimativas das correlações foram altas e positivas entre semanas adjacentes, com posterior decréscimo das mesmas com o aumento da distância entre os registros. As correlações genéticas (Figura 4), ao contrário do esperado em análises de regressão aleatória, apresentaram irregularidades. Dentre as irregularidades, verificou-se o surgimento de correlações genéticas negativas envolvendo as primeiras semanas de controle, sendo que no modelo BSP6 tais correlações antagônicas apresentam-se maiores. O declínio abrupto das correlações, envolvendo as primeiras semanas no modelo BSP6 em comparação ao modelo POL6, foram também constatados por Pereira et al. (2013) na raça Gir, sem entretanto ser verificado valor negativo.

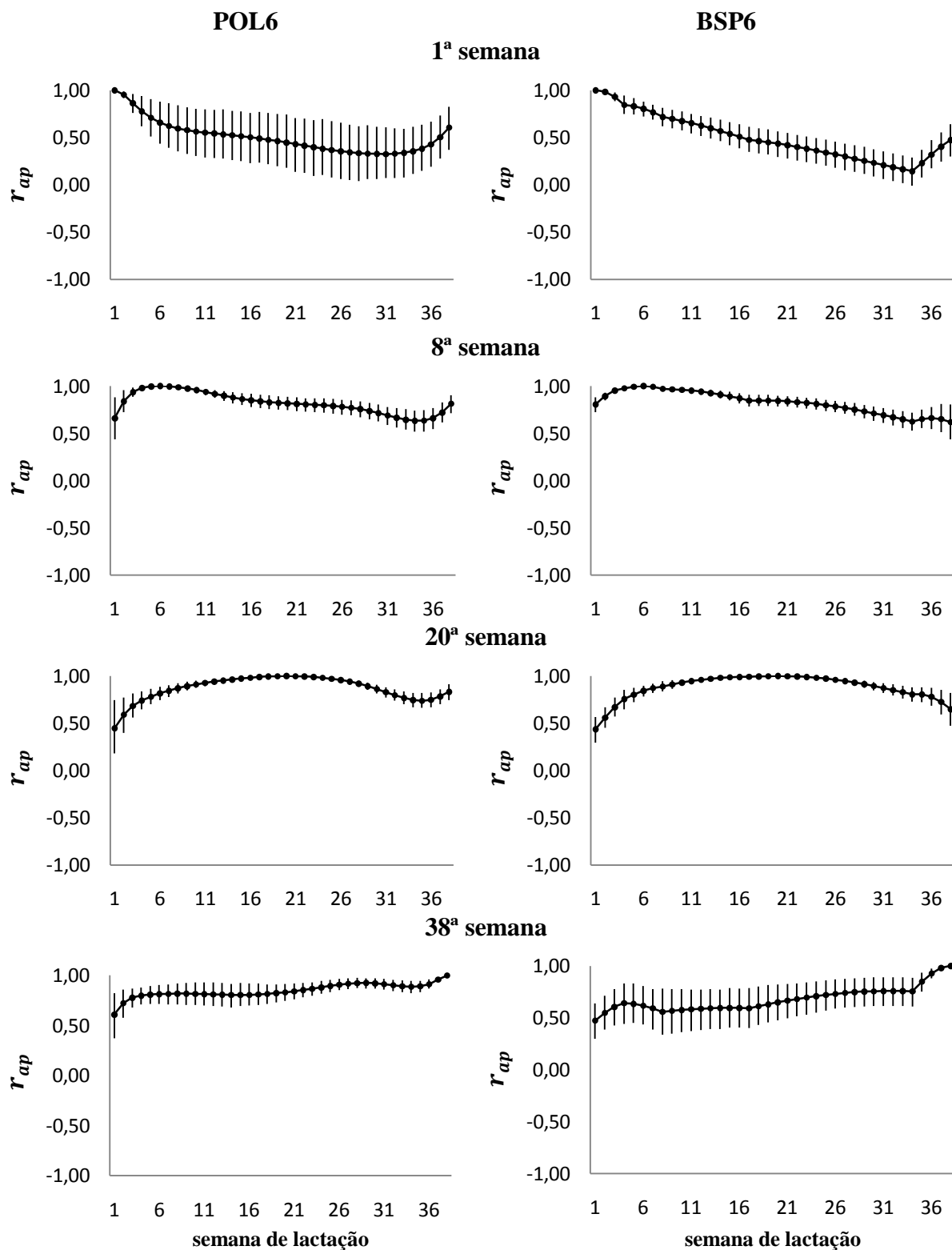


**Figura 4** – Médias (•) e 95% do intervalo de maior densidade posterior (barra) das distribuições posteriores das estimativas de correlações genéticas ( $r_g$ ) entre a 1ª, 8ª, 20ª e 38ª semana de lactação com as demais para os modelos POL6 e BSP6 em função da semana de lactação.

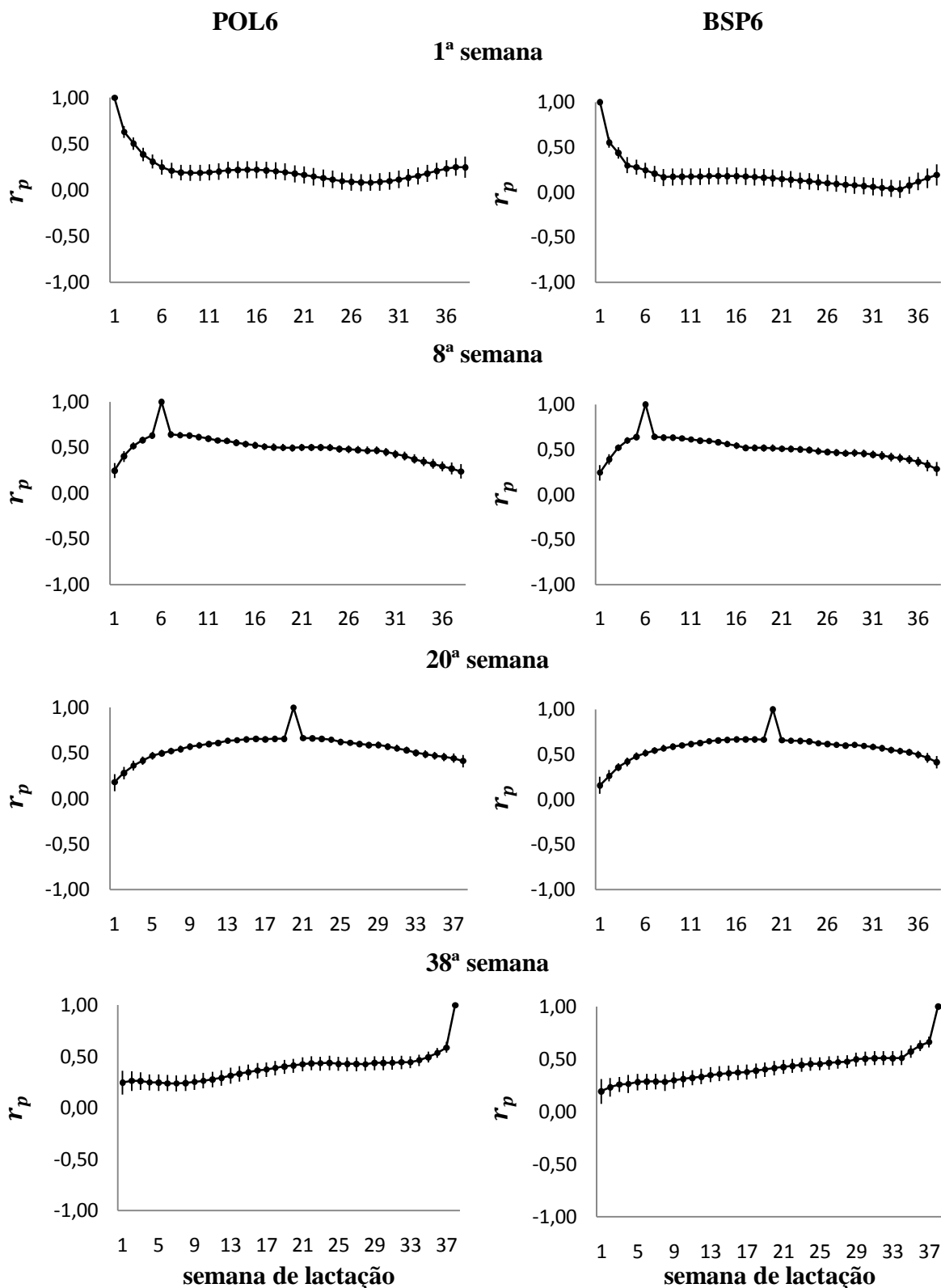
À exceção das primeiras semanas, a estrutura de correlação genética entre as demais semanas de lactação são próximas para ambos os modelos. Cabe ressaltar que o modelo utilizando polinômios de Legendre tendeu a apresentar um ajuste mais suave das curvas de correlação, fator desejável por traduzir uma realidade biológica de mudanças graduais na expressão gênica para as diferentes PLDC ao longo da lactação (PEREIRA, 2008).

Correlações genéticas negativas, assim como estruturas irregulares de correlação genética, são descritas em diversos estudos de regressão aleatória para produção de leite no dia do controle (EL FARO et al., 2008; BREDA et al., 2010; BIGNARDI et al., 2012). Sesana et al. (2010) atribuiu a ocorrência de estimativas negativas à influência do déficit energético da vaca no início da lactação e à dificuldade dos polinômios de Legendre em modelar registros nos extremos da curva. Bignardi et al. (2011) também encontrou estimativas negativas utilizando modelos de regressão aleatória com funções paramétricas (Ali e Schaeffer; Wilmink) e *B-splines*, de forma que pode-se inferir que independente da função base, tais estimativas podem ser oriundas de estruturas de dados deficientes (irregularidade na distribuição dos registros), da possibilidade de baixa conectividade entre os rebanhos, assim como de fatores ambientais não controlados que influenciam o início da lactação.

No que se refere às correlações de ambiente permanente e fenotípicas (Figuras 5 e 6), pode-se verificar que ambos os modelos apresentaram estruturas de correlações próximas e, assim como para o efeito genético, os polinômios de Legendre possibilitaram a construção de curvas mais suaves. Seus efeitos apresentam, entretanto, uma estrutura mais definida, onde as correlações tenderam a apresentar menores diferenças entre as estimativas à medida que a distância entre os controles aumentava. O efeito de ambiente permanente, como descrito em outros estudos de regressão aleatória com produções de leite no dia do controle (BREDA et al., 2010; SESANA et al., 2010; BIGNARDI et al., 2012; PEREIRA et al., 2013), apresentaram maiores correlações que os efeitos genéticos-aditivos - indicando que causas ambientais são as principais responsáveis pela semelhança no desempenho entre controles de um mesmo animal.



**Figura 5** – Médias (•) e 95% do intervalo de maior densidade posterior (barra) das distribuições posteriores das estimativas de correlações de ambiente permanente ( $r_{ap}$ ) entre a 1ª, 8ª, 20ª e 38ª semana de lactação com as demais para os modelos POL6 e BSP6 em função da semana de lactação.



**Figura 6** – Médias (•) e 95% do intervalo de maior densidade posterior (barra) das distribuições posteriores das estimativas de correlações fenotípicas ( $r_p$ ) entre a 1ª, 8ª, 20ª e 38ª semana de controle com as demais para os modelos POL6 e BSP6 em função da semana de lactação.

As estimativas de correlações, de uma forma geral, apresentaram intervalos de maior densidade posterior mais amplos que as de outros parâmetros sendo, portanto, necessária atenção às conclusões referentes a estes parâmetros. Os intervalos de maior densidade posterior relacionados aos primeiros controles apresentam uma grande amplitude comparados aos de outros controles, o que pode ser atribuído ao pequeno número de informações no atual período.

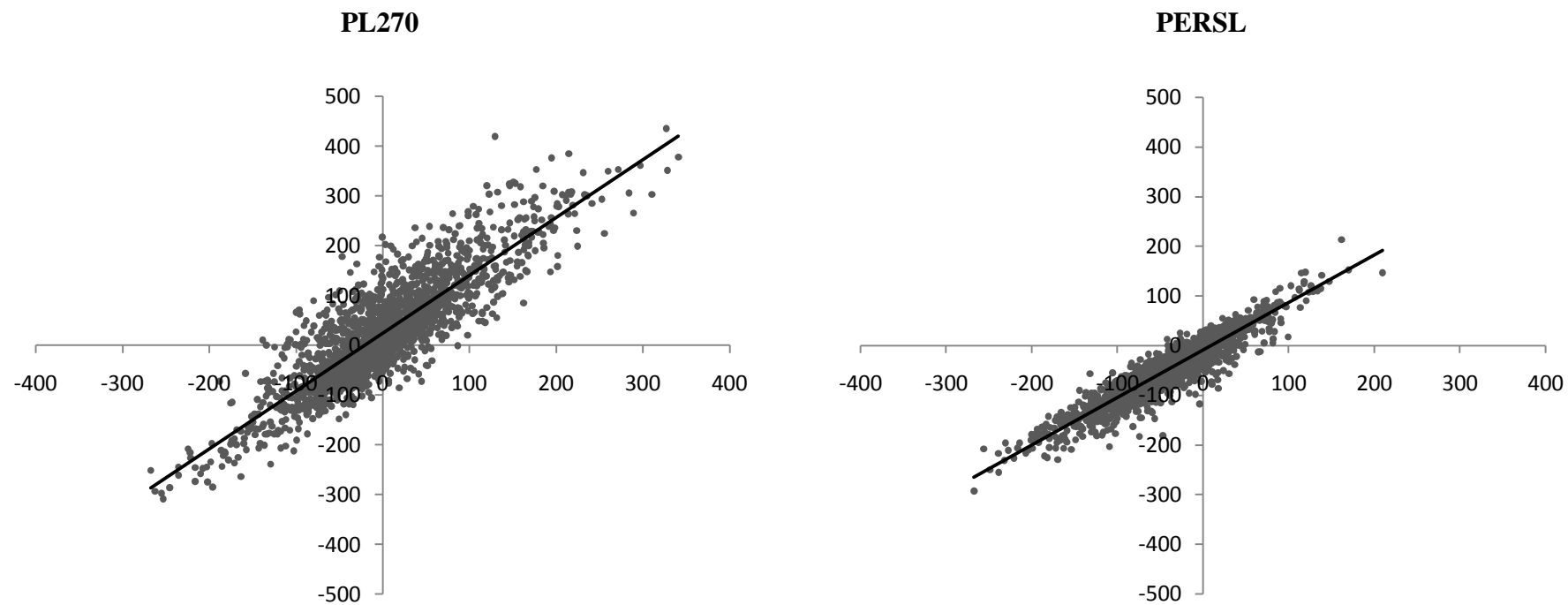
De uma forma geral, os resultados da utilização de modelos de regressão aleatória, utilizando *B-splines*, apresentaram resultados próximos aos do polinômio de Legendre, não sendo verificadas diferenças estatísticas entre muitos dos parâmetros estimados (verificados pela sobreposição dos intervalos de maior densidade posterior), à exceção da parte inicial.

São verificadas mudanças abruptas das curvas (variâncias, herdabilidade e correlações) no modelo BSP6 nos pontos referentes aos nós, o que de acordo com Misztal (2006) e Bohmanova et al. (2008) está relacionado à baixa correlação entre os nós, cujo valor deve se encontrar entre 0,6 e 0,8. Possivelmente pode-se associar também a isto a obtenção de estimativas anômalas obtidas, como a herdabilidade e correlações genéticas obtidas nos controles iniciais. Assim, a alternativa mais simples para minimizar o famigerado fenômeno, de acordo com Meyer (2005b) e Misztal (2006), seja alocar maiores números de nós nos pontos de mudança abrupta, tornando o modelo mais parametrizado e consequentemente mais suave. Alternativas adicionais serão discutidas posteriormente.

#### **5.4. Comparação dos valores genéticos (EBV): Produção de leite (PL270) e Persistência da lactação (PERSL)**

A associação entre os valores genéticos para produção de leite e persistência entre os modelos BSP6 e POL6 (Figura 7) apresentou-se alta e positiva, com valores de correlação de Pearson de 0,87 e 0,93 para PL270 e PERSL, respectivamente. Estes valores indicam que possivelmente a seleção dos melhores animais, baseados em um modelo específico, coincidirá na seleção dos melhores animais para outro modelo.

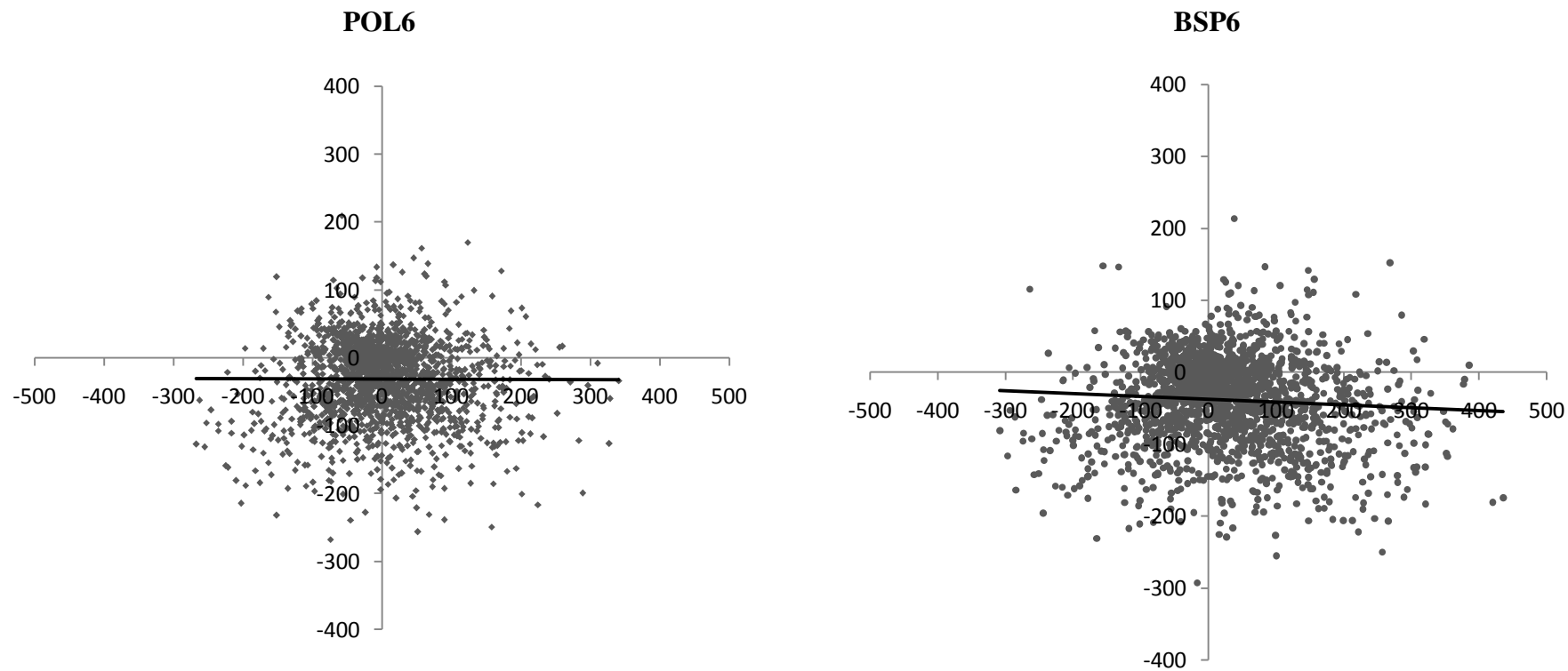
A maior associação entre os valores genéticos para PERSL pode ser atribuído ao fato de que para o cálculo de tal medida consideraram-se os períodos da lactação onde as estimativas de variâncias apresentaram maior similaridade entre POL6 e BSP6. Forte associação entre valores genéticos preditos para diferentes modelos de regressão (Legendre e *B-splines*) foram descritos por Santos (2011) e Pereira et al. (2013) para a raça Guzerá e Gir, respectivamente, com correlações de Pearson e Spearman superiores a 0,80.



**Figura 7** – Gráficos de dispersão entre valores genéticos estimados para os modelos POL6 e BSP6 para as características produção de leite acumulada aos 270 dias (PL270) e persistência de lactação (PERSL).

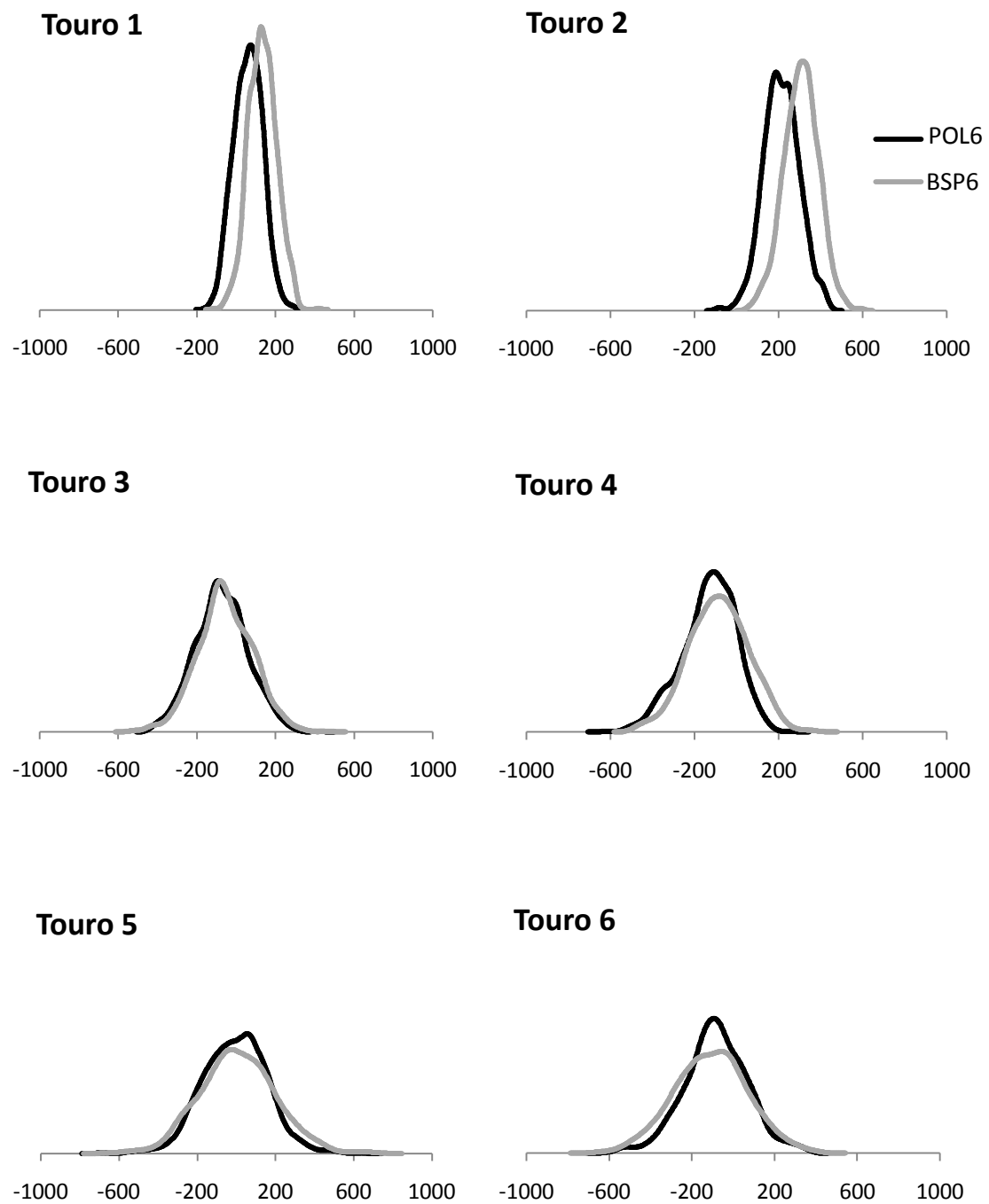
Pode-se também constatar em ambos os modelos que a associação entre os valores genéticos para PL270 e PERSL (Figura 8), cujos valores foram de -0,01 e -0,06, para POL6 e BSP6, respectivamente, apresentaram-se similares e de baixa associação. Segundo Dekkers et al. (1998) e Chaves (2009), medidas de persistência que apresentam correlações genéticas de baixas magnitudes com a produção de leite acumulada são as mais adequadas em processos de seleção, haja vista que será possível prover mudanças na curva de lactação sem infligir mudanças na produção de leite.

Os valores de correlação entre os valores genéticos para produção de leite aos 270 dias dos touros foram muitos similares aos encontrados para a população completa, seja considerando grupos de touros com pelo menos 1, 10 ou 30 filhos. Verificou-se também que ao se selecionar os 10 melhores animais para PL270, em cada um dos modelos, há 9 coincidentes - com algumas mudanças de posto (sendo o mesmo comportamento verificado para persistência da lactação).

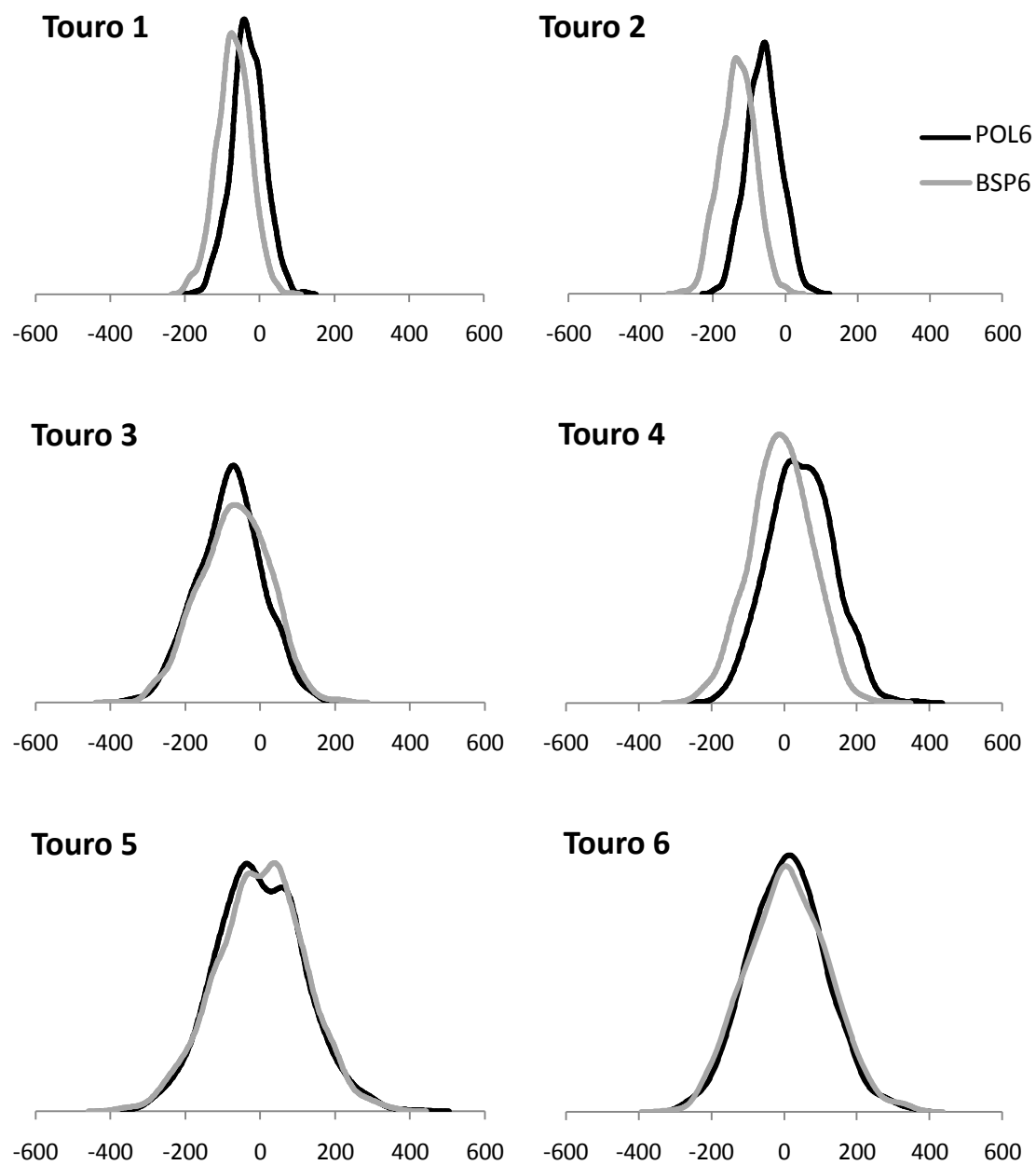


**Figura 8** – Gráficos de dispersão entre valores genéticos para produção de leite e persistência para os modelos POL6 e BSP6.

Com relação à habilidade de ambos os modelos em prever os valores genéticos para as características PL270 e PERSL de acordo com o número de filhas no rebanho, as distribuições posteriores dos valores genéticos para alguns touros que possuem mais de 60 progênes, 10 progênes e apenas uma progênie com dados de produção de leite encontram-se apresentadas nas Figuras 9 e 10. Primeiramente, verifica-se que, mesmo que os valores pontuais (média) dos valores genéticos sejam diferentes entre POL6 e BSP6, diferenças estatísticas não podem ser constatadas haja vista que há sobreposição dos intervalos de distribuição para qualquer número de filhas. Em segundo lugar, ambos os modelos devem apresentar valores similares de acurácia, visto que as distribuições dos valores genéticos apresentaram comportamento similar nas três situações apresentadas.



**Figura 9** – Densidades posteriores para os valores genéticos de produção de leite para os modelos POL6 (linha preta) e BSP6 (linha cinza) para touros com mais de 60 filhas (tousos 1 e 2), 10 filhas (tousos 3 e 4) e 1 filha (tousos 5 e 6).



**Figura 10** – Densidades posteriores para os valores genéticos de persistência de lactação para os modelos POL6 (linha preta) e BSP6 (linha cinza) para touros com mais de 60 filhas (touros 1 e 2), 10 filhas (touros 3 e 4) e 1 filha (touros 5 e 6).

## 6- CONCLUSÕES

Os modelos de regressão utilizados, polinômios ortogonais de Legendre e *B-splines*, podem ser indicados e utilizados satisfatoriamente para avaliação da produção de leite no dia de controle de bubalinos. Para banco de dados mais complexos, os modelos *B-splines* são mais indicados (dada sua facilidade de convergência).

## 7- PERSPECTIVAS FUTURAS

Em vista dos resultados apresentados, estudos futuros devem ser realizados em duas áreas de interesse:

a) efeito da estrutura dos dados sobre a estimação de parâmetros genéticos utilizando modelos de regressão aleatória e;

b) realização de estudos com base em diferentes algoritmos para definição de nós.

Majoritariamente, tem-se realizado estudos de estrutura utilizando dados reais, que dificultam o esclarecimento e associação de inúmeros efeitos ao comportamento das estimativas. Nos casos de dados simulados, os cenários consistem apenas de diferentes herdabilidades - em que podem ser inseridas e contabilizadas estruturas comumente observadas em populações reais, como o desbalanceamento no número de registros entre as classes de controle ao longo da curva de lactação, assim como o próprio tamanho da população.

Anomalias apresentadas em pontos da curva com menor número de registros demonstram a visível necessidade de estudos de simulação para comparar modelagens específicas, como *B-splines* e polinômios ortogonais de Legendre, conjuntamente a estudos de diferentes algoritmos para definição do número e posição dos nós para os primeiros, de modo que seja possível o emprego eficiente de *B-splines* no banco de dados de produção de leite em búfalas.

Com relação aos estudos de definição de nós pode-se, inicialmente, utilizar métodos que não necessitam modificar algoritmos de estimação de componentes de variâncias existentes. Ao fazer uso de modelos *B-splines*, Meyer (2005b) propõe a utilização da locação de nós equidistantes, de forma que a seleção do número seja baseada em critérios convencionais como BIC e AIC, sendo estas abordagens baseadas apenas no intervalo de tempo desejado.

Podem ser empregadas ainda metodologias que se baseiam no conhecimento da estrutura do banco de dados. Este procedimento foi adotado no presente estudo, pois a definição dos nós baseou-se na estrutura da curva de lactação da população. Conforme proposto por Misztal (2006), priorizou-se a alocação de nós em pontos de maior

concentração de observações, consequência da irregularidade das distribuições dos registros entre as classes de tempo. Na abordagem, as posições de alguns nós ditos principais são pré-definidos, sendo então os modelos testados nas análises diferenciados pela inserção de nós intermediários a estes. A presente metodologia pode ser aprimorada por meio do uso de testes para seleção de posições de *nós* candidatos, utilizando critérios de seleção como o DIC.

Métodos mais complexos têm sido desenvolvidos tanto na área de melhoramento animal como estatística, muitos em que há a necessidade modificar programas de estimação existentes. Jamrozik et al. (2010) propuseram a utilização de algoritmos genéticos sobre funções de (co)variâncias pré-estimadas (utilizando polinômios de Legendre), através do método de seleção diferencial, proposto por Price et al. (1997). O método facilita inclusive a seleção de nós de forma diferenciada para os efeitos genético-aditivos e de ambiente permanente, geralmente não empregado em diversos estudos com *splines* dada, principalmente, a quantidade de possibilidades de análises quando avaliado o número de nós, segmentos e ordens dos mesmos.

Outra possibilidade consiste em definir número e localização dos nós durante o processo de estimação de parâmetros, abordagem esta apresentada por Botts e Daniels (2008) e Kyung (2011) empregando métodos bayesianos em modelos lineares mistos. A dificuldade de tais abordagens consiste no fato de que grandes mudanças devem ser implementadas em programas existentes, pois utilizam não somente distribuições *a priori* específicas (como a distribuição Dirichlet para o segundo estudo), mas também esquemas de MCMC menos usuais (MCMC de saltos reversíveis para o primeiro estudo) - tornando necessário o desenvolvimento de novas sub-rotinas e programas. Entretanto, em vista das possibilidades da utilização dos *splines* na modelagem de dados longitudinais, torna-se interessante implementar tais metodologias em modelos de avaliação genética.

## 8- REFERÊNCIAS BIBLIOGRÁFICAS

- A.B.C.B. - Associação Brasileira dos Criadores de Búfalos. Disponível em: <<http://www.bufalo.com.br/>>. Acesso em: 12 dez. 2012
- ALBUQUERQUE, L.G. Regressão Aleatória: nova tecnologia pode melhorar a qualidade das avaliações genéticas. In: V SIMPÓSIO DA SOCIEDADE BRASILEIRA DE MELHORAMENTO ANIMAL, 2004, Pirassununga. **Anais...** Pirassununga: SBMA, 2004.
- ALI, T.E.; SCHAEFFER, R. Accounting for covariances among test day milk yield in dairy cows. **Canadian Journal of Animal Science**, n.67, p.637-644, set. 1987.
- ARAÚJO, C.V.; RAMOS, A.A.; ARAÚJO, S.I. et al. Buffalos milk yield analysis using random regression models. **Italian Journal of Animal Science**, v.6, n.2, p.279-282, 2007.
- ASPILCUETA-BORQUIS, R.R.; SESANA, R.C.; MUÑOZ-BERROCAL, M.; SENO, L.O.; BIGNARDI, A.B.; EL FARO, L.; ALBUQUERQUE, L.G.; CAMARGO, G.M.; TONHATI, H. Genetic parameters for milk, fat and protein yields in Murrah buffaloes (*Bubalus bubalis* Artiodactyla, Bovidae). **Genetics and Molecular Biology**, v.33, p.71-77, 2010.
- ASPILCUETA-BORQUIS, R.R.; ARAUJO NETO, F.R.; BALDI, F. et al. Genetic parameters for test-day yield of milk, fat and protein in buffaloes estimated by random regression models. **Journal of Dairy Research**, v.79, n.3, p.272-279, ago. 2012.
- BEAUMONT, M.A.; RANNALA, B. The Bayesian revolution in genetics. **Nature Reviews Genetics**, v.5, p.251-261, abr. 2004.
- BERNARDES, O. **Os búfalos no Brasil**. In: SIMPÓSIO DE BÚFALO DE LAS AMÉRICAS, 2., SIMPÓSIO EUROPA-AMÉRICA, 2., 2006, Medellín. **Anais...** Medellín. v.3, p.18-23, 2006.

- BIGNARDI, A.B.; EL FARO, L.; TORRES JR, R.A.A. et al. Random regression models using different functions to model test-day milk yield of Brazilian Holstein cows. **Genetics and Molecular Research**, v.10, n.4, p.3565-3575, 2011.
- BIGNARDI, A.B.; EL FARO, L.; SANTANA JR, M.L. et al. Bayesian analysis of random regression models using B-splines to model test-day milk yield of Holstein cattle in Brazil. **Livestock Science**, n.150, p.401-406, 2012.
- BLASCO A. The Bayesian controversy in animal breeding. **Journal of Animal Science**, v.79, n.8, p.2023-2046, 2001.
- BOHMANOVA, J., MIGLIOR, F., JAMROZIK, J. et al. Comparison of random regression models with Legendre polynomials and linear splines for production traits and somatic cell score of Canadian Holstein cows. **Journal of Dairy Science**, v.91, n.9, p.3627–3638, 2008.
- BOLIGON, A.A.; BALDI, F.; MERCADANTE, M.E.Z. Breeding value accuracy estimates for growth traits using random regression and multi-trait models in Nelore cattle. **Genetics and Molecular Research**, v.10, n.2, p.1227-1236, 2011.
- BOTTS, C.H.; DANIELS, M.J. A flexible approach to bayesian multiple curve fitting. **Computational Statistics & Data Analyses**, v.52, n.12, p.5100-5120, ago. 2008.
- BREDA, F.C.; ALBUQUERQUE, L.G.; EUCLYDES, R.F. et al. Estimation of genetic parameters for milk yield in Murrah buffaloes by Bayesian inference. **Journal of Dairy Science**, v.93, n.2, p.784-791, 2010.
- BUXADERA, A.M. **Principios básicos e importancia de lautilización de los sistemas de regresion aleatória em los programas de seleccion y mejora animal** (Apostila), Universidade de Córdoba, Espanha.
- CASELLA, G.; GEORGE, E.I. Explaining the Gibbs sampler. **The American Statistician**, v.46, n.3, p.167-174, ago. 1992.
- CHAVES, L.C.S. **Avaliação genética em bubalinos leiteiros utilizando modelos de regressão aleatória**. 2009. 66f. Tese (Doutorado em Zootecnia) – Universidade Federal de Viçosa, Minas Gerais, 2009.

- COBUCCI, J.A.; COSTA, C.N.; BRACCINI NETO, J. et al. Genetic parameters for Milk production by using random regression models with different alternatives of fixed regression modeling. **Revista Brasileira de Zootecnia**, v.40, n.3, p.557-567, 2011.
- DE BOOR, C. **A practical guide to splines**. Nova York: Springer-Verlag, v.27, 366 p., ed. rev., 2001.
- DEKKERS, J.C.M.; HAG, J.H.T.; WEERSINK, A. Economic aspects of persistency of lactation dairy cattle. **Livestock Production Science**, n.53, p.237-252, mar. 1998.
- DORNELES, C.K.P.; COBUCCI, J.A.; RORATO, P.R.N. et al. Estimação de parâmetros genéticos para produção de leite de vacas da raça Holandesa via regressão aleatória. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.61, n.2, p.407-412, 2009.
- DRUET, T.; JAFFREZIC, F.; DUCROQ, V. Estimation of genetic parameters for test day records of dairy traits in the first three lactations. **Genetics Selection Evolution**, v.37, p.257-271, 2005.
- EHLERS, R.S. **Introdução à inferência bayesiana** (Apostila). Curitiba: Universidade Federal do Paraná, 80 p., 5ª ed., 2007.
- EL FARO; L.; CARDOSO, V.L.; ALBUQUERQUE, L.G. Variance component estimates applying random regression models for test-day milk yield in Caracu heifers (*Bos taurus* Artiodactyla, Bovidae). **Genetics and Molecular Biology**, v.31, n.3, p.665-673, 2008.
- FAO - Food and Agriculture Organization. **Water buffalo: Regional Office for Asia and the Pacific**. Bangkok, Thailand, 2000.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.6, n.6, p.721-741, nov. 1984.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. **Bayesian Statistics 4**, Oxford: Oxford University Press, 31 p., 1992.

- GIANOLA, D.; FERNANDO, R.L. Bayesian methods in animal breeding theory. **Journal of Animal Science**, v.63, p.217-244, 1986.
- HALLANDER, J.; WALDMANN, P.; WANG, C. et al. Bayesian inference of genetic parameters based on conditional decompositions of multivariate normal distributions. **Genetics**, v.185, p.645-654, jun. 2010.
- HAMMAMI, H.; REKIK, B.; SOYEURT, H.; BEN GARA.; GENGLER, N. Genetic parameters for Tunisian Holsteins using a test-day random regression model. **Journal of Dairy Science**, v.91, n.5, p. 2118-2126, 2008.
- HEIDELBERGER, P.; WELCH, P.D. Simulation run length control in the presence of an initial transient. **Operations Research**, v.31, n.6, p.1109–1144, 1983.
- HENDERSON JR., C.R. Analysis of covariance in the mixed model: higher level, nonhomogeneous, and random regressions. **Biometrics**, v.38, n.3, p.623-640, set. 1982.
- HURTADO-LUGO, N.; CERÓN-MUÑOZ, M.; GUTIÉRREZ-VALENCIA, A. Estimación de parámetros genéticos para la producción de leche en el día de controle en búfalos de la Costa Atlántica Colombia. **Livestock Research for Rural Development**, v.18, n.3, 2006. Disponível em: <<http://lrrd.cipav.org.co/lrrd18/3/hurt18039.htm>>. Acesso em: 20 dez. 2013.
- IBGE – Instituto Brasileiro de Geografia e Estatística: Produção Pecuária Municipal, 2011. Disponível em: <<http://www.ibge.gov.br/home/estatistica/economia/ppm/2011/default.shtm>>. Acesso em: 12 jan. 2013.
- JAMROZIK, J.; SCHAEFFER, L.R.; DEKKERS, J.C.M. Genetic evaluation of dairy cattle using test day milk yields and random regression model. **Journal of Dairy Science**, n.80, v.6, p.1217-1226, 1997.
- JAMROZIK, J.; GIANOLA, D.; SCHAEFFER, L. R. Bayesian estimation of genetic parameters for test day records in dairy cattle using linear hierarchical models. **Livestock Production Science**, n.71, p.223–240, 2001.

- JAMROZIK, J. Implementation issues for Markov Chain Monte Carlo methods in random regression test-day models. **Journal of Animal Breeding and Genetics**, v.121, n.1, p.1-13, 2004.
- JAMROZIK, J.; BOHMANOVA, J.; SCHAEFFER, L. R. Selection of locations of knots for linear splines in random regression test-day models. **Journal of Animal Breeding and Genetics**, v.127, n.2, p.87-92, 2010.
- KIRKPATRICK, M.; LOFSVOLD, D.; BULMER, M. Analysis of the inheritance, selection and evolution of growth trajectories. **Genetics**, v.124, p.979-993, abr. 1990.
- KYUNG, M. A computational bayesian method for estimating the number of knots in regression splines. **Bayesian Analysis**, v.6, n.4, p.793-828, 2011.
- LÓPEZ-ROMERO, P.; REKAYA, R.; CARABAÑO, M.J. Bayesian comparison of test-day models under different assumptions of heterogeneity for the residual variance: the change point technique versus arbitrary intervals. **Journal of Animal Breeding and Genetics**, v.121, n.1, p.14-25, 2004.
- MENEZES, G.R.O.; TORRES, R.A.; SARMENTO, J.R.L. et al. Modelos de regressão aleatória na avaliação da produção de leite em cabras da raça Saanen. **Revista Brasileira de Zootecnia**, v.40, n.7, p.1526-1532, 2011.
- MEYER, K. Estimates of genetic covariances functions for growth of Angus cattle. **Journal of Animal Breeding and Genetics**, v.122, n.2, p.73-85, 2005a.
- MEYER, K. Random regression analyses using B-splines to model growth of Australian Angus cattle. **Genetic Selection Evolution**, v.37, p.473-500, 2005b.
- MISZTAL, I. Properties of random regression models using linear splines. **Journal of Animal Breeding and Genetics**, n.123, p.74-80. 2006.
- MISZTAL, I. 2012: BLUF90 family of programs. Disponível em: <ftp://num.ads.uga.edu/pub/blupf90/>. Acesso em: 18 dez. 2012.

- MURTEIRA, B.J.F. **Probabilidade e estatística: inferência estatística**. Lisboa: McGraw-Hill, v.2, 480 p., 1990.
- PENCHEV, P.; BOICHEV, M.; ILIEVA, Y. et al. Effect of different factors on lactation curve in buffalo cows. **Slovak Journal of Animal Science**, v.44, n.3, p.103-110, 2011.
- PEREIRA, J.C.C. **Melhoramento genético aplicado à produção animal**. Belo Horizonte: FEPMUZ Editora, 618 p., 2008.
- PEREIRA, R.J.; BIGNARDI, A.B.; EL FARO, L. et al. Random regression models using Legendre polynomials or linear splines for test-day milk yield of dairy Gyr (*Bos indicus*) cattle. **Journal of Dairy Science**, v.96, n.1, p.565-574, 2013.
- POOL, M.H.; JANSS, L.L.G.; MEUWISSEN, T.H.E. Genetic parameters of Legendre polynomials for first parity lactation curves. **Journal of Dairy Science**, v.83, n.11, p.2640–2649, 2000.
- PRICE K.V.; STORN, R.M.; LAMPINEN, J.A. **Differential evolution: A practical approach to Global optimization**. Nova York: Springer, 1997.
- R Development Core Team (2010). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- RESENDE, M.D.V.; REZENDE, G.D.S.P.; FERNANDES, J.S.C. Regressão aleatória e funções de covariância na análise de medidas repetidas. **Revista de Matemática e Estatística**, v.19, p.21-40, 2001.
- RODRIGUES, C.F.C.; IAPICHINI, J.E.C.B.; LISERRE, A.M. et al. Oportunidades e desafios da bubalinocultura familiar da região sudoeste paulista. **Revista tecnologia & Inovação Agropecuária**, p.100-109, dez. 2008.
- SANTIAGO, A.A. **Introdução dos Búfalos no Brasil**. Associação Brasileira dos Criadores de Búfalos, 13p., 2000.

- SANTOS, D.J.A. **Diferentes abordagens para modelar a produção de leite de bovinos da raça Guzerá**. 2011. 141f. Dissertação (Mestrado em Genética e Melhoramento Animal) – UNESP/FCAV/Câmpus de Jaboticabal – São Paulo, 2011.
- SAS Institute Inc. **Statistical Analysis System user's guide**. Version 9.1 ed. Cary: SAS Institute, USA, 2003.
- SCHAEFFER, L.R. Application of random regression models in animal breeding. **Livestock Production Science**, v.86, n.1, p.35-45, 2004.
- SCHUMAKER, L.L. **Spline functions: basic theory**. Nova York: Cambridge University Press, 3<sup>a</sup> ed., 2007.
- SESANA, R.C.; BIGNARDI, A.B.; ASPILCUETA-BORQUIS, R.R. et al. Random regression models to estimate genetic parameters for test-day milk yield in Brazilian Murrah buffaloes. **Journal of Animal Breeding and Genetics**, v.127, n.5, p.369-376, 2010.
- SILVA, F.G. **Estudo da curva de produção de leite de caprinos utilizando modelos de regressão aleatória**. 2011. 144f. Dissertação (Mestrado em Genética e Melhoramento) – Universidade Federal de Viçosa, Minas Gerais, 2011.
- SILVA, M.M.A.; BARROS, N.A.M.T.; RANGEL, A.H.N. et al. Persistência da lactação em búfalas da raça Murrah (*Bubalus bubalis*) exploradas no agreste do Rio Grande do Norte. **Acta Veterinaria Brasilica**, v.4, n.4, p.286-293, 2010.
- SMITH, B.J. BOA: An R Package for MCMC Output Convergence Assessment and Posterior Inference. **Journal of Statistical Software**, v.21, n.11, p.1-37, nov., 2007.
- SORENSEN, D.; GIANOLA, D. **Likelihood, Bayesian, and MCMC methods in quantitative genetics**. Nova York: Springer, 740p., 2002.
- SPIEGELHALTER, D.J.; BEST, N.G.; CARLIN, B.P. et al. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society, Series B. Statistical Methodology**, v.64, p.583-639, 2002.

- STRABEL, T., MISZTAL, I. Genetic parameters for first and second lactation milk yields of Polish Black and White cattle with random regression test-day models. **Journal of Dairy Science**, v.82, n.12, p.2805-2810, 1999.
- STRABEL, T.; JAMROZIK, J. Genetic analysis of milk production traits of Polish Black and White cattle using large-scale random regression test-day models. **Journal of Dairy Science**, v.89, n.8, p.3152-3163, 2006.
- TONHATI H.; MUÑOZ, M.F.C.; DUARTE J.M.C. et al. Estimates of correction factors for lactation length and genetic parameters for milk yield in buffaloes. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, v.56, n.2, 251–257, abr. 2004.
- TONHATI, H.; MENDOZA-SÁNCHEZ, G.; SESANA, R.C. et al. Programa de mejoramiento genético de búfalos. **Revista Argentina de Producción Animal**, v.28, n.1, p.53-67, 2008.
- TONHATI, H.; SESANA, R.C.; ALBUQUERQUE, L.G. Avaliação genética de búfalos leiteiros. In: **Avaliação genética de búfalos leiteiros** (TONHATI, H. e ALBUQUERQUE, L.G., ed.). Fundação de Apoio a Pesquisa, Ensino e Extensão – Funep, Jaboticabal, 101 p., 2007.
- VAN DER WERF, J. **Random regression in animal breeding**. Course notes. Jaboticabal, Brasil, nov. 2001. Disponível em: <<http://wwwpersonal.une.edu.au/~jvanderw/CFcoursenotes.pdf>>. Acesso em: 20 dez. 2012.
- VAN TASSELL, C.P.; VAN VLECK, L.D. **A manual for use of MTGSAM. A set of Fortran programs to apply Gibbs Sampling to animal models for variance components estimation**. Lincoln: Department of Agriculture, Agricultural Research Service, 85p., 1995.
- WHITE I.M.S.; THOMPSON, R.; BROTHERSTONE S. Genetic and environmental smoothing of lactation curves with cubic splines. **Journal of Dairy Science**, v.82, n.3, p.632–638, 1999.