**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Parametrização de Sistemas de Equações Diferenciais
Ordinárias no crescimento de bovinos de corte e produção de
gases**

Adriele Giaretta Biase

Tese apresentada para obtenção do título de Doutora
em Ciências. Área de concentração: Estatística e Ex-
perimentação Agronômica

**Piracicaba
2015**

Adriele Giaretta Biase
Matemática

**Parametrização de Sistemas de Equações Diferenciais Ordinárias no crescimento de
bovinos de corte e produção de gases**

Orientador:
Prof. Dr. **CARLOS TADEU DOS SANTOS DIAS**

Tese apresentada para obtenção do título de Doutora em Ciências.
Área de concentração: Estatística e Experimentação Agronômica

**Piracicaba
2015**

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Biase, Adriele Giaretta

Parametrização de Sistemas de Equações Diferenciais Ordinárias no crescimento de bovinos de corte e produção de gases / Adriele Giaretta Biase. - - Piracicaba, 2015.
146 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Otimização 2. Pacote FME 3. *Bootstrap* não-paramétrico 4. Incerteza do parâmetro
I. Título

CDD 636.213
B579p

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

AGRADECIMENTO

A Deus, que sempre iluminou e guiou a minha vida, dando-me forças para vencer e concretizar os meus sonhos.

À Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES) e à Empresa Brasileira de Pesquisa Agropecuária (Embrapa) pelo apoio financeiro nos estudos e na pesquisa que tornou a realização desse trabalho possível, em especial ao Projeto PECUS-PC7 - Bioma Amazônia. Deixo meus sinceros agradecimentos às instituições ESALQ/USP e Texas A&M University por todas as condições oferecidas para a minha formação.

Ao orientador Carlos Tadeu dos Santos Dias e amigos Luís Gustavo Barioni, Lucieta Guerreiro Martorano e Luis Orlindo Tedeschi, por todos os conselhos, pela atenção e pelas palavras sábias que sempre me incentivaram e iluminaram meus caminhos. A dedicação e profissionalismo serão minha inspiração por toda a vida.

Aos meus pais, José Luiz e Elisa, pelo infinito amor e por terem lutado incansavelmente pela minha educação e que nunca mediram esforços para que meus objetivos fossem alcançados.

Ao Fernando, pela paciência, por me escutar, por ser meu grande amigo e companheiro de todas as horas e sempre me apoiou em todas as minhas decisões de minha vida. Obrigada pelo amor incondicional, você é meu alicerce.

Às minhas irmãs, Érica e Nádia, por sempre caminharem ao meu lado. Aos meus cunhados, que por diversas vezes foram tão prestativos.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP, pelos ensinamentos, pela amizade e excelente convivência.

Aos colegas do curso de Pós-graduação em Estatística e Experimentação Agronômica. Foi um imenso prazer conhecê-los e conviver com todos durante esse tempo.

A todos os funcionários do Departamento de Ciências Exatas, em especial ao Eduardo, Jorge, Luciane, Mayara, Rosni e Solange (técnicos auxiliares e secretárias da Pós-graduação) pela prontidão e suporte.

Ao Tiago Zanetti Albertini, principalmente pela amizade, disponibilidade e entusiasmo transmitidos durante todos esses anos de doutorado. Muito obrigada!

Às amigas feitas no Texas, no Department of animal science: Samantha, Whitney, Jackson Kirby, Jocelyn, Abby Cupples, Elizabeth, Vinícius e Mozart pelo carinho e me ajudarem a enriquecer meus conhecimentos; Aos amigos Mônica, George Lyerly, Leanne, Georginho e Andrea pela acolhida e momentos inesquecíveis durante essa caminhada.

Às amigas no pensionato e “Cenáculo N. Sr^a de Lurdes” que me proporcionaram momentos de alegria e diversão na casa de *n* mulheres. Nele foram vários “Encontros e Partidas”.

Aos meus sobrinhos, mesmo que pequeninos, foram tão maravilhosos com os sorrisos inocentes.

Muito obrigada a todos!

EPÍGRAFE

“ A estrada para o sucesso não é uma reta,
há uma curva chamada fracasso,
um trevo chamado confusão,
um quebra-molas chamado obstáculos,
faróis de advertência chamados retornos e
pneus furados chamados dificuldades.
Mas se tiver um estepe chamado determinação,
um motor chamado perseverança,
um seguro chamado fé e
um motorista chamado Deus,
você chegará a um lugar chamado
sucesso...”

Pe. Marcelo Rossi

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE TABELAS	13
LISTA DE FIGURAS	17
1 INTRODUÇÃO	21
1.1 Referencial teórico	23
1.1.1 Inferência <i>Bayesiana</i>	23
1.1.2 <i>Bootstrap</i>	29
1.1.3 Séries Temporais	31
Referências	39
2 COMPARAÇÃO DO AJUSTE DE EQUAÇÕES: UM ESTUDO DE PRODUÇÃO DE GÁS <i>IN VITRO</i>	43
Resumo	43
Abstract	43
2.1 Introdução	44
2.2 Material e Métodos	46
2.2.1 Equação não-linear na forma analítica (integral exata)	46
2.2.2 Equação diferencial ordinária	48
2.2.3 Banco de dados	49
2.2.4 Processos numéricos iterativos: algoritmos de otimização	50
2.2.5 <i>Bootstrap</i> não-paramétrico	52
2.2.6 Simulação Monte Carlo via Cadeia de Markov	52
2.2.7 Análise de sensibilidade e Colinearidade	56
2.2.8 Avaliação de modelos	57
2.2.9 Teste de aleatorização de Mantel	60
2.2.10 Análise de Séries Temporais	61
2.3 Resultados e Discussões	62
2.3.1 Banco de dados sintéticos	62
2.3.2 Conjunto de dados baseados na produção de gás <i>CO₂ in vitro</i>	65
2.4 Conclusões	84
Agradecimentos	84
Referências	85
3 MODELO MATEMÁTICO DINÂMICO PARA ESTIMAR INDIVIDUALMENTE O DESEMPENHO DE BOVINOS DE CORTE CRUZADOS	91
Resumo	91
Abstract	91
3.1 Introdução	92

3.2 Material e Métodos	92
3.3 Resultados e Discussões	99
3.4 Conclusões	106
Agradecimentos	106
Referências	106
4 PARAMETRIZAÇÃO DO MODELO DE CRESCIMENTO DE DAVIS PARA BOVINOS DE CORTE CRUZADOS	109
Resumo	109
Abstract	109
4.1 Introdução	110
4.2 Material e métodos	112
4.2.1 Base de dados	112
4.2.2 Modelo	112
4.2.3 Estatística Computacional	112
4.2.4 Avaliação de modelos	116
4.3 Resultados e Discussões	116
4.3.1 Comparações da abordagem baseadas nos estudos individuais e na amostra total . .	125
4.4 Conclusões	127
4.5 Trabalhos futuros	127
Agradecimentos	128
Referências	128
ANEXO	133

RESUMO

Parametrização de Sistemas de Equações Diferenciais Ordinárias no crescimento de bovinos de corte e produção de gases

Parametrizações de modelos e estruturas de correlações dos parâmetros no âmbito agropecuário são importantes por caracterizarem o comportamento de um sistema em resposta a variações de múltiplos cenários (clima, genótipos, dietas nutricionais, dentre outros fatores) que existem em escalas globais. O objetivo foi contribuir com inferências estatísticas na produção de gases CO_2 [um potente Gás de Efeito Estufa (GEE)] nas fermentações *in vitro* de feno de alfafa, comparando métodos frequentistas com novas metodologias surgidas na literatura científica como a combinação dos métodos de Rejeição por Atraso e o Metrópolis Adaptativo (RAMA), até então não testados para predições de gases de fermentação *in vitro*. Além disso, modelos de séries temporais foram usados para previsão da produção de CO_2 nas fermentações de gases *in vitro* de feno de alfafa. Dentro do contexto de crescimento de gado de corte, foi realizada pela primeira vez uma abordagem para predições individuais dos animais para taxa de ganho de peso e a necessidade de energia para manutenção baseada na dinâmica de crescimento e composição química corporal do Modelo de Crescimento de Davis (MCD), com comparação de análise de covariância multivariada entre diferentes cenários (gêneros, sistemas e genótipo cruzados), em um experimento a campo no Brasil. Adicionalmente calibrações dos parâmetros baseadas na amostra de cada cenário, pelos ajustes do MCD e usando análise frequentista, *bootstrap* não-paramétrico e simulações Monte Carlo foram realizadas com os dados nacionais (raça cruzada) e comparada com as estimativas originais do modelo obtido com raças Britânicas (*Bos taurus*). Os principais critérios adotados para avaliar os ajustes dos modelos foram o Erro Quadrático Médio de Predição (EQMP), o Critério de Informação Akaike (AIC) e o Critério de Informação Bayesiano (BIC). Os resultados não só contribuirão para o avanço da literatura existente, mas também auxiliarão a indústria de carne bovina e produtores rurais a encontrar especificações do mercado de carne, tanto a nível nacional e internacional. Concluiu-se que *i*) na produção de gases: o modelo ARIMA (1, 1, 2) ajustou a produção acumulativa de CO_2 , atingindo o valor máximo de 1,1066 (mL) no tempo de 47,5 h e a equação é indicada para estimar a produção de gases; *ii*) no crescimento de gados de corte usando as estimativas individuais do MCD, os vetores de efeitos de energia de manutenção e o acréscimo de proteína possuem efeitos pronunciados quanto as interações entre sistemas e gêneros; *iii*) no crescimento de gados de corte usando as estimativas da amostra total com MCD, os genótipos cruzados tiveram maior gasto de energia de manutenção e foram mais rápidos de maturação em comparação tanto com os animais de genótipos Britânicos (*Bos Taurus*) e touros Nelores. A técnica de *bootstrap* não-paramétrica estimou com sucesso as distribuições dos parâmetros (que tiveram distribuição probabilidade normal para maioria dos cenários). Correlação negativa entre os parâmetros de acréscimo de DNA e energia de manutenção foram encontrados para animais machos não castrados do sistema extensivo, indicando que foram mais eficientes no uso da energia. A generalização de tal relação ainda demanda estudos mais abrangentes e aprofundados.

Palavras-chave: Otimização; Pacote FME; *Bootstrap* não-paramétrico; Incerteza do parâmetro

ABSTRACT

Parameterization of Ordinary Differential Equations Systems in the growth of beef cattle and production of gases

Model parameter fitting and parameter correlation structures are important for characterize a system's behaviour in response to multiple scenarios variations (climate, genotypes, nutritional diet and other factors). The aim was to contribute to statistical inferences in the production of CO_2 [a potent greenhouse gas (GHG)] *in vitro* fermentation of alfalfa hay, comparing frequentist methods with new methodologies that emerged in the scientific literature, such as the combination of a delay Rejection and the Adaptive Metropolis methods (RAMA), not yet tested for *in vitro* fermentation gases predictions. In addition, time series models were used to predict CO_2 production in the *in vitro* fermentation of alfalfa hay. For the first time, individual predictions of animal weight gain rate and energy of maintenance based on the growth dynamics and body composition Davis Growth Model (DGM) was carried out besides multivariate covariance analysis of different scenarios (genres, systems and crossed genotype). Additionally, parameter estimation based on sample of each scenario, using frequentist analysis, nonparametric bootstrap and Monte Carlo simulations were performed with national data (cross breed) and compared to the original estimates of the model obtained with British breeds (*Bos taurus*). The main criteria used to evaluate the model accuracy were the Mean Square Error of Prediction (MSEP), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The results not only contribute to the scientific progress in modeling animal growth and composition, but also assist the beef industry and farmers to adjust the production process to the meat market specifications, both nationally and internationally. For *in vitro* gas production, we concluded that the ARIMA (1, 1, 2) model presented the highest accuracy in predicting cumulative CO_2 production and the respective equation is recommended for estimating CO_2 production. In the growth beef cattle using the individual estimates DGM, average vectors from maintenance of energy and protein accretion showed pronounced effects as the interactions between systems and genres. Also the total sample with DGM, cross-breed genotypes had higher maintenance energy expenditure and were faster-maturing compared with British genotypes animals (*Bos Taurus*) and Nellore bulls estimates. Bootstrap nonparametric with downhill simplex optimization method successfully estimated the distributions of the parameters (that had normal probability distribution for most scenarios). Uncastrated male animals of the extensive system showed negative correlation between the protein deposition rate and requirement for energy maintenance, indicating that animals with faster lean tissue deposition were also more efficient in energy usage. We warn that the generalization of this finding demands studies with larger populations.

Keywords: Optimization; FME package; Nonparametric bootstrap; Uncertainty parameter

LISTA DE TABELAS

Tabela 2.1 -	Simulação do banco de dados sintético para comparar inferências estatísticas usando a forma analítica e EDO com solução frequentista e MCMC . . .	50
Tabela 2.2 -	Informações <i>prioris</i> usadas e <i>posteriores</i> geradas pelos modelos não-lineares por meio da metodologia bayesiana - produção de gás CO_2 de ruminantes <i>in vitro</i>	56
Tabela 2.3 -	Fator de dependência do critério de Raftery e Lewis (FD) e valor p do critério de Geweke dos modelos não-lineares - dados sintético	63
Tabela 2.4 -	Resultados da comparação dos modelos não-lineares por análise frequentista e MCMC na forma analítica - <i>priori</i> não informativa - banco de dados sintético	63
Tabela 2.5 -	Resultados da comparação dos modelos não-lineares por análise frequentista e MCMC na forma EDO - <i>priori</i> não informativa - banco de dados sintético	64
Tabela 2.6 -	Avaliação de modelos não-lineares metodologias frequentista e MCMC (<i>priori</i> não informativa) - dados sintéticos	65
Tabela 2.7 -	Intervalos de confiança para as médias dos parâmetros com 95% de confiança e teste Kolmogorov-Smirnov (KS) para os modelos não-lineares por análise <i>bootstrap</i> com 2.000 reamostragens - produção de gases CO_2 com fermentação <i>in vitro</i>	66
Tabela 2.8 -	Avaliação dos modelos não-lineares para estimativas de emissões de gases CO_2 no período de 48 horas para feno de alfafa - análise das estimativas de <i>bootstrap</i> com 2.000 reamostragens	67
Tabela 2.9 -	Fator de dependência do critério de Raftery e Lewis (FD) e valor p do critério de Geweke dos modelos não-lineares - produção de gases CO_2 <i>in vitro</i>	68
Tabela 2.10 -	Resultados da comparação dos modelos não-lineares por análise frequentista e MCMC na forma analítica e EDO - <i>prioris</i> normais - produção de gases CO_2 com fermentação <i>in vitro</i>	68
Tabela 2.11 -	Avaliação dos modelos não-lineares para estimativas de emissões de gases CO_2 no período de 48 horas para feno de alfafa - forma analítica	69
Tabela 2.12 -	Avaliação dos modelos não-lineares para estimativas de emissões de gases CO_2 no período de 48 horas para feno de alfafa - EDO	69
Tabela 2.13 -	Artigos acadêmicos que retratam produção de gases de efeito estufa na agropecuária no Brasil com publicações desde o ano 2000	74
Tabela 2.14 -	Teste aleatorização de Mantel para comparar as matrizes de correlações dos parâmetros dos modelos estimados pela metodologia <i>bootstrap</i> e MCMC, com 1.000 permutações	77
Tabela 2.15 -	Estimativas dos parâmetros dos modelos de séries temporais ajustados aos gases CO_2 em mL , para o período de 48 horas	80

Tabela 2.16 -	Estatísticas para o teste de Box-Pierce de modelos ARIMA	82
Tabela 2.17 -	Estimativas dos critérios de AIC, BIC e EQMP de modelos ARIMA	82
Tabela 2.18 -	Valores observados de gases CO_2 , para o período de $t = 47,5$ à $t = 48$ horas e valores preditos para o mesmo período	83
Tabela 3.1 -	Resumo com as análise descritiva [médias, desvio padrão (DP), mínimo (Min) e máximo (Max)] das variáveis de entrada do modelo - Fêmeas	94
Tabela 3.2 -	Resumo com as análise descritiva [médias, desvio padrão (DP), mínimo (Min) e máximo (Max)] das variáveis de entrada do modelo - Machos	95
Tabela 3.3 -	Análise de covariância univariada considerando a taxa de ganho de peso (k_1)	102
Tabela 3.4 -	Comparação das médias de taxas de acréscimo de proteína (fêmeas sistema extensivo, $n = 4$; fêmeas sistema intensivo, $n = 9$; machos sistema extensivo, $n = 5$; machos sistema intensivo, $n = 9$) e teste de Tukey	103
Tabela 3.5 -	Análise de covariância univariada considerando a exigência de energia para manutenção (α)	103
Tabela 3.6 -	Comparação de médias de energia para manutenção (fêmeas sistema extensivo, $n = 4$; fêmeas sistema intensivo, $n = 9$; machos sistema extensivo, $n = 5$; machos sistema intensivo, $n = 9$) e Teste de Tukey	103
Tabela 3.7 -	Análise de covariância multivariada considerando os parâmetros k_1 e α	105
Tabela 3.8 -	Distâncias quadráticas de <i>Mahalanobis</i> e valores descritivos de probabilidade pelo teste F para os contrastes multivariados, considerando as combinações entre os níveis do gênero e os níveis do sistema	105
Tabela 4.1 -	Estimativas e Intervalos de confiança frequentistas e <i>bootstrap</i> não-paramétrico para o MCD	117
Tabela 4.2 -	Medidas descritivas, teste Kolmogorov-Smirnov (valor p) e intervalos de confiança para as médias dos parâmetros com 95% de confiança baseadas nas Simulações Monte Carlo	119
Tabela 4.3 -	Avaliação do MCD com as estimativas obtidas pelas análises frequentistas, <i>bootstrap</i> e simulação Monte Carlo para quatro cenários de estudos (fêmeas e machos nos sistemas extensivos e intensivos)	120
Tabela 4.4 -	Comparações das estimativas médias de proteína e gordura entre as novas calibrações propostas nesse estudo e as calibrações usando os valores de Oltjen et al. (1986)	121
Tabela 4.5 -	Porcentagem de mudança dos atuais valores dos parâmetros calibrados para genótipos cruzados (por <i>Bootstrap</i>) em relação aos valores dos parâmetros originais do MCD com genótipo Britânico <i>Bos taurus</i> realizado por Oltjen et al. (1986) e com a parametrização de Nelore <i>Bos Indicus</i> realizado por Sainz et al. (2006)	123

Tabela 4.6 -	Análise residual para cada compartimento das variáveis de estado do MCD expresso pelo valor p dos testes Breusch-Pagan, Durbin-Watson e Shapiro-Wilk, baseadas na estimativas de <i>bootstrap</i> não-paramétrico	124
Tabela 4.7 -	Análise residual do MCD expresso pelo valor p dos testes Breusch-Pagan, Durbin-Watson e Shapiro-Wilk, baseadas na estimativas de <i>bootstrap</i> não-paramétrico	124
Tabela 4.8 -	Comparações das hipóteses baseados nos ajustes individuais e nos ajustes baseados na amostra total baseado no MCD, Oltjen et al. (1986), ao nível de 5% de significância	125

LISTA DE FIGURAS

Figura 2.1 -	Diagrama com a composição metodológica da proposta desse estudo	46
Figura 2.2 -	Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 , θ_3 , θ_4 e θ_5) na forma EDO Logístico bicompartimental - produção de gases CO_2 fermentação de feno de alfafa experimento <i>in vitro</i>	70
Figura 2.3 -	Ajuste MCMC da equação não-linear Logística bicompartimental - produção de gases (CO_2) fermentação de feno de alfafa experimento <i>in vitro</i> . . .	71
Figura 2.4 -	Sensibilidades dos parâmetros do Logístico bicompartimental - produção de gases (CO_2) fermentação de feno de alfafa experimento <i>in vitro</i>	71
Figura 2.5 -	Matriz de dispersão conjunta com os histogramas de frequência dos ajustes <i>bootstrap</i> (à esquerda) e ajustes MCMC (à direita) para produção de gases <i>in vitro</i> no período de 48 horas, modelo exponencial na forma analítica . . .	75
Figura 2.6 -	Matriz de dispersão conjunta com os histogramas de frequência dos ajustes <i>bootstrap</i> (à esquerda) e ajustes MCMC (à direita) para produção de gases <i>in vitro</i> no período de 48 horas, equação logística na forma analítica	75
Figura 2.7 -	Matriz de dispersão conjunta com os histogramas de frequência dos ajustes <i>bootstrap</i> (à esquerda) e ajustes MCMC (à direita) para produção de gases <i>in vitro</i> no período de 48 horas, equação logística na forma analítica	76
Figura 2.8 -	Matriz de dispersão conjunta com os histogramas de frequência dos ajustes <i>bootstrap</i> (à esquerda) e ajustes MCMC (à direita) para produção de gases <i>in vitro</i> no período de 48 horas, equação logística na forma de EDO	76
Figura 2.9 -	Representação gráfica da série original de produção de gases no período de 47,4167 horas, em mL	77
Figura 2.10 -	Periodograma da série de produção de gases no período de 47,4167 horas, em mL	78
Figura 2.11 -	Representação gráfica da série diferenciada de produção de gases no período de 47,4167 horas, em mL	79
Figura 2.12 -	Função de autocorrelação e autocorrelação parcial da série diferenciada de produção de gases no período de 47,4167 horas, em mL	79
Figura 2.13 -	Função de autocorrelação residual e autocorrelação parcial residual da série temporal, modelo ARIMA (1, 1, 1)	81
Figura 2.14 -	Função de autocorrelação residual e autocorrelação parcial residual da série temporal, modelo ARIMA (1, 1, 2)	81
Figura 2.15 -	Função de autocorrelação residual e autocorrelação parcial residual da série temporal, modelo ARIMA (3, 1, 0)	81
Figura 2.16 -	Função de autocorrelação residual e autocorrelação parcial residual da série temporal, modelo ARIMA (4, 1, 0)	82
Figura 2.17 -	Previsão da produção de gases CO_2 <i>in vitro</i> após o período de 47,5 horas . .	84

Figura 3.1 -	Análise de covariância - dispersão das idades (dias) em relação aos parâmetros k_1 (a) e α (b)	100
Figura 3.3 -	Interação entre gênero e sistema para a variável k_1	100
Figura 3.2 -	Análise de covariância - dispersão do peso inicial (kg) em relação aos parâmetros k_1 (a) e α (b)	101
Figura 3.4 -	Interação entre gênero e sistema para a variável α	101
Figura 3.5 -	Elipses de 95% de confiança dos vetores de médias para as fêmeas no sistema intensivo, fêmeas no sistema extensivo, machos no sistema intensivo e machos no sistema extensivo de acordo com as variáveis k_1 transformada [potência ótima $\lambda = -0,6000$ (Box-Cox, 1964)] e α	106
Figura 4.1 -	Dispersão conjunta, correlação de <i>Pearson</i> e histograma de frequência para k_1 e α para o MCD por meio da análise de <i>bootstrap</i> não-paramétrico . . .	118
Figura 4.2 -	Dispersão conjunta, correlação de <i>Pearson</i> e histograma de frequência para k_1 e α para o MCD por meio de simulações Monte Carlo	118
Figura 4.3 -	Valores observados e estimados para as proteínas e as gorduras - modelo DGM ajustando com os parâmetros estimados da proposta desse estudo (à direita) e com os parâmetros originais de referência de touros <i>Bos Taurus</i> (à esquerda)	122
Figura 4.4 -	Dispersão dos resíduos estudantizados em relação a ordem das observações para os cenários do experimento parametrização <i>bootstrap</i> não-paramétrico do MCD	124
Figura 4.5 -	Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$ e $p3 = \theta_3$) estimados por MCMC, modelo analítico exponencial com perturbação $N(0; 0, 05^2)$	135
Figura 4.6 -	Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 , θ_3) e variância (σ^2) do modelo analítico exponencial com perturbação $N(0; 0, 05^2)$	135
Figura 4.7 -	Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$ e $p3 = \theta_3$) estimados por MCMC, modelo exponencial com perturbação $N(0; 0, 2^2)$	136
Figura 4.8 -	Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) do modelo analítico exponencial com perturbação $N(0; 0, 2^2)$	136
Figura 4.9 -	Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$ e $p3 = \theta_3$) estimados por MCMC, modelo analítico logístico com perturbação $N(0; 0, 05^2)$	137
Figura 4.10 -	Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) do modelo analítico logístico com perturbação $N(0; 0, 05^2)$	137

Figura 4.11 - Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$ e $p3 = \theta_3$) estimados por MCMC, modelo logístico com perturbação $N(0; 0, 2^2)$	138
Figura 4.12 - Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) do modelo analítico logístico com perturbação $N(0; 0, 2^2)$	138
Figura 4.13 - Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$, $p3 = \theta_3$, $p4 = \beta_4$ e $p5 = \beta_5$) estimados por MCMC, modelo analítico logístico bicompartimental com perturbação $N(0; 0, 05^2)$	139
Figura 4.14 - Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 , θ_3 , θ_4 e θ_5) e variância (σ^2) do modelo analítico logístico bicompartimental com perturbação $N(0, 0, 05^2)$	139
Figura 4.15 - Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$, $p3 = \theta_3$, $p4 = \beta_4$ e $p5 = \beta_5$) estimados por MCMC, modelo analítico logístico bicompartimental com perturbação $N(0; 0, 2^2)$	140
Figura 4.16 - Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 , θ_3 , θ_4 e θ_5) e variância (σ^2) do modelo analítico logístico bicompartimental com perturbação $N(0; 0, 2^2)$	140
Figura 4.17 - Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($a = \theta_1$, $b = \theta_2$ e $c = \theta_3$) estimados da EDO exponencial, MCMC, com perturbação $N(0; 0, 05^2)$	141
Figura 4.18 - Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) com EDO exponencial com perturbação $N(0; 0, 05^2)$	141
Figura 4.19 - Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($a = \theta_1$, $b = \theta_2$ e $c = \theta_3$) estimados por MCMC, com EDO exponencial com perturbação $N(0; 0, 2^2)$	142
Figura 4.20 - Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) do modelo analítico exponencial com perturbação $N(0; 0, 2^2)$	142
Figura 4.21 - Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($a = \theta_1$, $b = \theta_2$ e $c = \theta_3$) estimados da EDO logístico, MCMC, com perturbação $N(0; 0, 05^2)$	143
Figura 4.22 - Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) com EDO logístico com perturbação $N(0; 0, 05^2)$	143
Figura 4.23 - Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($a = \theta_1$, $b = \theta_2$ e $c = \theta_3$) estimados da EDO logístico, MCMC, com perturbação $N(0; 0, 2^2)$	144
Figura 4.24 - Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) com EDO logístico com perturbação $N(0; 0, 2^2)$	144

Figura 4.25 -	Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($a = \theta_1$, $b = \theta_2$, $c = \theta_3$, $d = \theta_4$ e $e = \theta_5$) estimados da EDO logístico bicompartimental, MCMC, com perturbação $N(0; 0, 05^2)$	145
Figura 4.26 -	Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 , θ_3 , θ_4 e θ_5) e variância (σ^2) com EDO logístico bicompartimental com perturbação $N(0; 0, 05^2)$	145
Figura 4.27 -	Matriz de dispersão, histograma de frequência e correlações dos parâmetros estimados da EDO logístico bicompartimental, MCMC, com perturbação $N(0; 0, 2^2)$	146
Figura 4.28 -	Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 , θ_3 , θ_4 e θ_5) e variância (σ^2) com EDO logístico bicompartimental com perturbação $N(0; 0, 2^2)$	146

1 INTRODUÇÃO

A caracterização do comportamento dos intrincados sistemas biológicos nos quais se dá a produção agropecuária exige a correta determinação do valor dos parâmetros e de sua estrutura de correlações. Modelos ajustados a contextos regionais específicos garantem maior previsibilidade e podem ser interessante instrumento para adaptação de sistemas para maior lucratividade e menor impacto ambiental.

O sistema de equações diferenciais proposto por Oltjen et al. [1986, denominado de Modelo de Crescimento de Davis (MCD)] para estimar a composição corporal de gado da raça Britânica (*Bos taurus*) vem sendo usado pelo programa de predições fenotípicas da “Cooperative Research Centre” (CRC) e tem se mostrado promissor em pesquisas nos Estados Unidos (McPhee et al., 2007; Walmsley et al., 2014).

Em razão das diferenças de padrão de crescimento entre raças de bovinos, bem como em função da grande diversidade de sistemas de produção animal existentes no Brasil, a reparametrização do MCD não foi realizada para as condições tropicais, especificamente no Brasil para genótipos provenientes de cruzamentos (*Bos indicus* x *Bos taurus*). Prever o crescimento desses animais por meio de modelos dinâmicos possibilita simular os efeitos da intensidade de restrição e realimentação sobre a manutenção, composição do ganho e a ingestão durante intervalos de tempo. Nesse sentido, a modelagem matemática é essencial para imitar o funcionamento de um sistema real nas estimativas de parâmetros de modelos de crescimento sob o ponto de vista estatístico e biológico. Além disso, podem ampliar sua abrangência a um custo mais baixo e tempo mais curto que a experimentação (Barioni et al., 2011). Ademais, as simulações por meio de modelos matemáticos podem evitar prejuízos por evitar decisões equivocadas e otimizar a decisão e o funcionamento do sistema real.

Modelos matemáticos de crescimento de bovinos de corte são potencialmente melhores quando contemplam o metabolismo e a disponibilidade de nutrientes. Assim, dentre esses modelos mais complexos, que englobam variáveis de estado e sistemas de equações diferenciais ou de diferença, destaca-se os modelos formalizados por Oltjen et al. (1986), France et al. (1987), Williams e Bennett (1995), Williams e Jenkins (2003), Hoch e Agabriel (2004) e Tedeschi et al. (2004). Esses modelos são capazes de representar o sistema real, porém, são de implementação e interpretação mais difíceis que os modelos estáticos tradicionais em virtude de se caracterizarem por sistema de equações e possuírem geralmente maior número de parâmetros.

Entre os modelos dinâmicos citados na literatura científica mencionada, o modelo do Oltjen et al. (1986) é um dos mais parcimoniosos com apenas cinco parâmetros interpretáveis. Além disso, esse modelo é baseado em conceitos de hiperplasia e hipertrofia do animal, que advêm da utilização de variáveis de estado como a massa de proteína, massa de DNA e massa de gordura que representam melhor o sistema de crescimento de bovinos. Essas características contempladas no modelo garantem maior precisão da previsão e aplicação ao longo de um leque mais vasto de condições que modelos derivados empiricamente (OLTJEN et al., 1986).

Em se tratando de impactos ambientais na produção animal, a emissão de metano pela fermentação entérica foi apontada como uma fonte de gases com efeito substancial na escala global. A pecuária contribui com cerca de 18% do total mundial as emissões antrópicas de Gases de Efeito Estufa (GEE), segundo Steinfeld et al. (2006). A utilização de modelos não-lineares podem reduzir os erros associados com a previsão de metano e permitir uma melhor análise e representação das políticas que regulam as emissões de gado. Entidades governamentais vem procurando quantificar as emissões de metano que se baseiam nesses modelos (Moraes et al., 2013).

A disponibilidade de informações sobre os sistemas de produção de gases emitidos por bovinos aumentou vastamente em função de técnicas como a incubação *in vitro* com líquido ruminal, para determinar o padrão de fermentação dos alimentos, principalmente devido a sua facilidade de utilização e custo (Tedeschi, 2008). Nesse sentido, esse trabalho buscou também contribuir com inferências estatísticas na produção de gases, comparando métodos frequentistas com metodologias que tem recebido, recentemente, grande atenção na literatura científica, por meio de Monte Carlo via Cadeia de Markov (MCMC), combinando dois métodos a Rejeição por Atraso e o Metrópolis Adaptativo (RAMA), até então não utilizado para previsões de gases de fermentação *in vitro*.

Inferência estatística bayesiana fornece uma ferramenta poderosa para modelar sistemas de equações não-lineares na agropecuária. Métodos MCMC fornecem formas de realizar cálculos do modelo de forma unificada, prático e intuitivamente plausível. Em contraste com a análise estatística clássica, análises envolvendo *bootstrap* e bayesiana fornecem uma maneira eficaz para o estudo das incertezas. Essas análises podem lidar com estimativas de parâmetros correlacionados e até mesmo, a análise bayesiana, pode diagnosticar matrizes de covariância singulares.

O método RAMA, implementado no pacote FME [*software* free R, Soetaert e Petzoldt (2010)] para geração de realizações de variáveis aleatórias com a estrutura de uma matriz de covariância multivariada dos parâmetros tem mostrado resultados promissores para estimação de parâmetros em sistemas de EDO e equações não-lineares. Estudos com simulações comparando convergências do método de Metropolis-Hastings e o método RAMA, mostram que esse último método fornece um diagnóstico sistemático para remediar o processo de adaptação da cadeia, Haario et al. (2006).

O método de *bootstrap* é utilizado para estimar características de interesse como: viés, variância, quantis ou a distribuição de amostragem do estimador. Trata-se de uma ideia conceitualmente muito simples mas as aplicações práticas desta metodologia têm-se revelado de grande importância quando se usa estatísticas para as quais a distribuição de amostragem é ou desconhecida ou intratável, Efron e Tibshirani (1993).

Deste modo, metodologias como de análise bayesiana e *bootstrap* não-paramétrico vem sendo aplicado em diversas áreas para obter a distribuição assintótica dos parâmetros e inferências a partir desses resultados. Essas metodologias corroboram para a melhoria da precisão e

a acurácia suprindo situações como as de pequenas amostras, por exemplo, elas são comumente usadas nas áreas médicas e sistemas econômicos (Fei, 2014; Ugarte et al., 2014; Zapf et al., 2015). Essas metodologias são promitentes para na aplicação de sistemas complexos do campo agropecuário.

No Capítulo 2 uma abordagem envolvendo equações não-lineares, exponencial, logística e logística bicompartimental foram ajustadas usando banco de dados sintético e de fermentação *in vitro* envolvendo degradação de feno de alfafa. Essas equações não-lineares foram comparadas usando análises frequentistas e bayesianas (RAMA) quanto a predição da produção de gases CO_2 . Adicionalmente, modelos de séries temporais como ARIMA (auto-regressivo integrado de médias móveis) também foram ajustados para contemplar a dependência residual da produção acumulativa da produção de gases CO_2 .

No Capítulo 3 uma abordagem baseada nos ajustes individuais do MCD foram realizadas num banco de dados de bovinos de corte provenientes de cruzamento racial para estimar das taxas de acréscimo de DNA e energia de manutenção do MCD. Análises de covariâncias univariadas e multivariadas foram realizadas posteriormente para comparar a dinâmica de crescimento e composição química corporal dos cenários do experimento a campo no Brasil, envolvendo gêneros e sistemas.

No Capítulo 4 uma abordagem baseada na soma de quadrados residuais da amostra total de cada cenário foi realizada com o objetivo de estimar os parâmetros do MCD, para representar lotes homogêneos de indivíduos, usando o mesmo banco de dados de genótipo cruzado do experimento a campo no Brasil (Capítulo 3). Calibrações dos parâmetros do MCD usando a análise frequentista, *bootstrap* não-paramétrico e simulações Monte Carlo foram realizadas com os dados nacionais, animais cruzados (*Bos taurus* × *Bos indicus*) e comparada com as estimativas originais do modelo obtido com raças Britânicas (*Bos taurus*), Oltjen et al. (1986).

1.1 Referencial teórico

1.1.1 Inferência Bayesiana

A metodologia bayesiana consiste de informações referentes aos dados amostrais (função de verossimilhança), do conhecimento prévio a respeito dos parâmetros (*priori*) e da obtenção da densidade a posteriori dos parâmetros por meio do teorema de Bayes. A utilização de informação a *priori* requer a especificação de uma distribuição para o parâmetro, devendo esta representar probabilisticamente o conhecimento que se tem sobre ele, antes da realização do experimento. A diferença formal entre a inferência *bayesiana* e a frequentista é que na inferência *bayesiana*, o parâmetro θ é uma variável aleatória, possuindo, então, uma distribuição de probabilidade. Na inferência frequentista, os parâmetros são valores fixos, não sendo possível atribuir a eles uma distribuição de probabilidade (Toral et al., 2007). Já a variável medida é atribuída uma distribuição de probabilidade.

Teorema de Bayes

Se $\pi(\boldsymbol{\theta})$ é a densidade *a priori* de $\boldsymbol{\theta}$, então, a densidade *a posteriori* de $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta}|\mathbf{Y})$, é dada pelo Teorema de Bayes:

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) = \frac{\pi(\mathbf{Y}, \boldsymbol{\theta})}{\pi(\mathbf{Y})} = \frac{L(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (1.1)$$

em que $\mathbf{Y} = \{y_1, y_2, y_3, \dots, y_n\}$. Note que $\frac{1}{\pi(\mathbf{Y})}$ funciona como uma constante normalizadora já que não depende de $\boldsymbol{\theta}$.

Assim, o teorema pode ser reescrito como:

$$\pi(\boldsymbol{\theta}|\mathbf{Y}) \propto L(\mathbf{Y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \quad (1.2)$$

sendo que $\propto$ representa a proporcionalidade. Em outras palavras a expressão (1.2) pode ser entendida como:

$$\text{distribuição a posteriori} \propto \text{Verossimilhança} \times \text{priori} \quad (1.3)$$

No caso de $\boldsymbol{\theta}$ ser multivariado, $(\boldsymbol{\theta} = \theta_1, \theta_2, \dots, \theta_p)$, as distribuições marginais das componentes θ_i , a partir das quais as inferências para cada parâmetro são feitas, podem ser obtidas da densidade conjunta *a posteriori* $\pi(\theta_1, \theta_2, \dots, \theta_p|y_1, y_2, \dots, y_p)$.

A densidade marginal *a posteriori* de θ_i é dado por

$$\pi(\theta_i|\mathbf{Y}) = \int \pi(\theta_1, \theta_2, \dots, \theta_p|\mathbf{Y})d\boldsymbol{\theta}_{-i} \quad (1.4)$$

em que $\boldsymbol{\theta}_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$ é o vetor $\boldsymbol{\theta}$ com a sua i -ésima componente removida.

A distribuição condicional completa do parâmetro θ_i (denotada por $\pi(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{Y})$) é obtida considerando que, na densidade conjunta, os demais parâmetros ($\boldsymbol{\theta}_{-i}$) são conhecidos e assim, a expressão se torna menos complexa, já que as constantes podem ser desconsideradas.

Quando a expressão da condicional completa tem a forma de uma densidade conhecida e portanto, fácil de ser amostrada, um método de simulação indicado é o Amostrador de Gibbs, um processo iterativo que gera valores que convergem para a densidade marginal, sem que se conheça a sua expressão.

Se a distribuição condicional completa *a posteriori* não é uma densidade conhecida, outros métodos de simulação são indicados como a amostragem por importância (PAULINO et al., 2003), a amostragem por aceitação e rejeição (CHIB E GREENBERG, 1995) e o algoritmo de *Metropolis-Hastings* (HASTINGS, 1970).

Métodos Monte Carlo via Cadeias de Markov

i) *Algoritmo de Gibbs.*

1. definem-se os valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$ para os parâmetros;
2. amostra-se iterativamente

$$\begin{aligned} \theta_1^{(1)} & \text{ de } \pi_1(\theta_1 | \theta_2^{(0)}, \dots, \theta_p^{(0)}, \mathbf{Y}) \\ \theta_2^{(1)} & \text{ de } \pi_2(\theta_2 | \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_p^{(0)}, \mathbf{Y}) \\ & \vdots \\ \theta_p^{(1)} & \text{ de } \pi_p(\theta_p | \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{p-1}^{(1)}, \mathbf{Y}) \end{aligned}$$
 obtendo-se na primeira iteração $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_p^{(1)})$.
3. repete-se o passo (2.) q vezes, ou seja, após um grande número de iterações (q iterações), obtém-se $\theta^{(q)} = (\theta_1^{(q)}, \theta_2^{(q)}, \dots, \theta_p^{(q)})$

O conjunto dos q valores amostrados representam amostras da distribuição marginal *a posteriori* de θ . A partir dessa distribuição obtêm-se as estimativas pontuais (médias, medianas e modas) e estimativas por região para os parâmetros amostrados.

ii) *Metropolis-Hastings*

O algoritmo *Metropolis-Hastings* é utilizado quando o amostrador de *Gibbs* não se mostra eficiente, ou seja, quando a distribuição condicional de θ não se caracteriza como uma distribuição de probabilidade conhecida.

O método consiste em gerar candidatos (θ^*) de uma densidade auxiliar $q(\theta, \theta^*)$ que possa ser amostrada e rejeitar ou aceitar esses valores com probabilidade de aceitação

$$\alpha(\theta, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^*)q(\theta^*, \theta)}{\pi(\theta)q(\theta, \theta^*)} \right\} \quad (1.5)$$

Descrição do algoritmo de Metropolis-Hastings:

1. inicialize o contador de iterações $j = 0$ e especifique um valor inicial $\theta^{(0)}$.
2. gerar um valor $\theta^{(*)}$ da distribuição $q(\theta^{(j)}, .)$;
3. calcular $\alpha(\theta, \theta^{(*)})$ e gerar $u \sim U(0, 1)$;
4. se $u \leq \alpha$, então aceite o valor de θ e faça $\theta^{(j+1)} = \theta^{(*)}$. Caso contrário, $\theta^{(j+1)} = \theta^{(j)}$
5. mudar o contador de j para $j + 1$ e voltar ao passo (2.) até a convergência.

iii) *Metropolis Adaptativo - MA*

O algoritmo MA foi proposto por Haario et al. (2001). Segundo esses autores a sintonização da matriz de covariância da distribuição *posteriori* θ pode ser simulada com sucesso usando o algoritmo de Metropolis Hasting e aplicando uma proposta com distribuição Gausiana. A sincronização dos novos candidatos propostos com distribuição Gausiana multivariada dado a estrutura de correlação dos parâmetros, segundo Haario et al. (2001) é importante para casos de simulações de modelos altamente não-lineares, quando existe correlação entre os componentes da parte *posteriori* ou quando a quantidade de parâmetro é alta.

1. Comece de um valor inicial θ^0 e uma matriz de covariância proposta $C = C_0$. Selecione um fator de escala s e um número ε para a regularização da covariância e um período n_0 de não-adaptação inicial;
2. Para cada passo, uma nova proposta θ^* de uma distribuição Gausiana centrada no atual valor $N(\theta^{i-1}, C)$ é gerada;
3. Aceita ou rejeita θ^* de acordo com a probabilidade de aceitação do algoritmo de Metropolis Hasting;
4. Após um período inicial de simulação, $i \geq n_0$, a atualização da matriz de covariância é feita usando a cadeia gerada até o momento por $C = cov(\theta^0, \dots, \theta^i)s + I\varepsilon$

A atualização pode ser feita em intervalos fixos ou aleatórios.

5. Iterativas simulações são feitas de (2) até valores suficientes serem gerados.

iv) *Rejeição por atraso - RA*

No algoritmo RA, proposto por Mira (2001), propõe que após a rejeição de um ponto candidato proposto, ao invés do próximo candidato proposto ser avaliado com a mesma taxa de probabilidade de aceitação, um segundo passo é proposto. A probabilidade de aceitação do candidato da segunda fase é calculada de modo que a reversibilidade da cadeia de Markov em relação à distribuição de interesse seja preservada. O processo de retardar a rejeição pode ser repetido para qualquer número fixo ou aleatório de vezes e as propostas dos estágios mais elevados são permitidas e dependem dos candidatos até o momento propostos e rejeitados. Assim a RA permite a adaptação local parcial da proposta dentro de cada passo de tempo da cadeia de Markov, ainda mantendo a propriedade Markoviana e reversibilidade.

O primeiro estágio da probabilidade de aceitação na RA é realizada de acordo com o algoritmo de Metropolis-Hasting padrão, que pode ser escrito como

$$\alpha_1(\theta, \theta^*) = \min \left\{ 1, \frac{\pi(\theta^*)q_1(\theta^*, \theta)}{\pi(\theta)q_1(\theta, \theta^*)} \right\} \quad (1.6)$$

em que θ é o ponto atual, θ^* é o novo valor proposto retirado a partir da distribuição de densidade auxiliar $q_1(\theta, \cdot)$ e da distribuição alvo π . Se θ^* for rejeitado, um segundo candidato θ^{**} é retirado de $q_2(\theta, \theta^*, \cdot)$ usando a probabilidade de aceitação

$$\alpha_2(\theta, \theta^*, \theta^{**}) = \min \left\{ 1, \frac{\pi(\theta^{**})q_1(\theta^{**}, \theta^*)q_2(\theta^{**}, \theta^*, \theta)[1 - \alpha_1(\theta^{**}, \theta^*)]}{\pi(\theta)q_1(\theta, \theta^*)q_2(\theta, \theta^*, \theta^{**})[1 - \alpha_1(\theta, \theta^*)]} \right\}. \quad (1.7)$$

Como a reversibilidade é preservada esse algoritmo também leva a mesma distribuição estacionária π como o algoritmo de Metropolis-Hasting, Mira (2001).

v) Combinação da Rejeição por atraso e Metropolis Adaptativo- RAMA

1. Comece de um valor inicial θ^0 e uma matriz de covariância proposta $C^{(1)} = C_0$ para a primeira fase. Selecione um fator de escala s , um número ε para a regularização da covariância, um período n_0 de não-adaptação inicial e um fator de escala para os estágios mais elevados propostos por $C^{(i)}$, $i = 1, \dots, ntrydr$, em que $ntrydr$ é o número de tentativas permitidas.
2. Realize o *Loop* DR, até que um novo valor é aceito, ou tentativas $ntrydr$ foram feitas:
 - (a) Gerar θ^* de uma distribuição Gaussiana centrada no valor atual $N(\theta^{i-1}, C^{(k)})$.
 - (b) Aceitar de acordo com a probabilidade de aceitação do k -ésimo estágio.
3. Definir $\theta^i = \theta^*$ ou $\theta^i = \theta^{i-1}$ de acordo se aceita-se ou não o valor proposto.
4. Depois de um inicial período de simulação $i \geq n_0$, adaptar-se a covariância proposta usando a cadeia gerada até agora, definida como: $C^{(1)} = cov(\theta^0, \dots, \theta^i)s + I\varepsilon$.

Calcule a proposta dos estágios mais elevados como versões em escala de $C^{(1)}$, de acordo com a regra escolhida.

5. Iterativas simulações do passo (2) em diante são realizadas, até que valores suficientes forem gerados.

Nos métodos bayesianos não existem um critério de parada como no caso dos algoritmos frequentistas. No caso do algoritmo de Gibbs, a cadeia tem 100% de aceitação. Para os demais algoritmos bayesianos descritos anteriormente, a aceitação de um novo valor proposto para o parâmetro dependem das probabilidades de aceitação [por exemplo: Eqs. (1.5) e (1.7)].

Critério de Convergência

Os algoritmos MCMC são processos iterativos e as cadeias resultantes necessitam ter a sua convergência constatada. Os métodos ou critérios utilizados para diagnosticar, ou seja,

para monitorar a convergência, podem ser classificados como informais e formais. Os métodos informais são as técnicas gráficas para a verificação da convergência, como, por exemplo, observar a trajetória da cadeia ao longo das iterações. Se após um período inicial, apresentar repetidamente o mesmo comportamento, então, conclui-se que houve convergência.

Para avaliação da convergência desses algoritmos existem também vários testes formais como o testes de diagnósticos de Geweke (1992), Gelman e Rubin (1992), Raftery-Lewis e o de Heidelberg e Welch (1983). Esses testes podem ser aplicados usando o pacote *coda*, Plummer et al. (2006), do *software* livre R (R Development Core Team, 2014).

i) Critério de Raftery e Lewis

O critério proposto por Raftery e Lewis (1992) é um método que estima quantas iterações são necessárias para que o amostrador de Gibbs apresente convergência à distribuição estacionária, propondo quantidades necessárias para isso. Ao se analisar a convergência de uma sequência é comum descartar as primeiras iterações, em geral, de 40% a 50% do total (Gamerman, 1997), considerando-se que essa primeira parte esteja sendo influenciada pelos valores iniciais. Este início da cadeia é chamado de período de “aquecimento”.

Outro aspecto importante refere-se à dependência entre as observações subsequentes da cadeia. Para se obter uma amostra independente, as observações devem ser espaçadas por um determinado número de iterações, ou seja, considerar saltos de tamanho k , usando, para compor a amostra, os valores a cada k iterações.

O critério de Raftery e Lewis fornece estimativas do número de iterações necessárias para se obter a convergência, do número de iterações iniciais que devem ser descartadas e da distância mínima (k) de uma iteração à outra para se obter uma amostra independente. Esses valores são calculados mediante especificações para garantir que um quantil u de uma determinada função $f(\theta)$ seja estimado com uma precisão predefinida.

A regra de decisão do critério de Raftery e Lewis é baseada no Fator de Dependência (FD). De acordo com Raftery e Lewis (1992) se o FD for maior que 5, pode-se concluir que a cadeia não atingiu convergência.

ii) Critério de Geweke

O critério proposto por Geweke (1992) consiste em indicar a convergência da média a posteriori da quantidade amostrada, $g(\theta)$, em que uma única cadeia, isto é, os valores $g(\theta^{(i)})$ são calculados a cada interação da simulação, formando uma série temporal. A partir da cadeia, a variância assintótica $S^2(0)$ da medida de $g(\theta)$ pode ser estimada. O critério de Geweke consiste nos seguintes passos:

Passo 1: dividir a cadeia de tamanho N em duas sequências. A primeira sequência possui N_a primeiras iterações e a segunda N_b restantes iterações. Geweke (1992) sugere que as

médias devem ser estimadas após o descarte de algumas iterações iniciais e que sejam considerados $N_a = 0,1 N$ e $N_b = 0,5 N$.

Passo 2: estimam-se as médias $\hat{g}_a$ e $\hat{g}_b$ e as variâncias assintóticas $\hat{S}_{g_a}^2(0)$ e $\hat{S}_{g_b}^2(0)$ de cada sequência, respectivamente. As variâncias são determinadas pela estimação da densidade espectral.

O critério garante que, se as razões N_a/n e N_b/n são fixas, quando $n \rightarrow \infty$, então:

$$\frac{\hat{g}_a - \hat{g}_b}{\sqrt{\left(\hat{S}_{g_a}^2(0)/N_a\right) + \left(\hat{S}_{g_b}^2(0)/N_b\right)}} \sim N(0, 1).$$

Portanto, um teste pode ser construído e se a diferença padronizada entre as médias for muito grande, existe indicação de ausência de convergência. A regra de decisão do critério de Geweke é baseada na estimativa do valor p . De acordo com Geweke (1992), se o valor p for menor que um nível de significância adotado pelo pesquisador, conclui-se que a cadeia não atingiu convergência.

1.1.2 *Bootstrap*

O *bootstrap* foi introduzido por Efron no final da década de 1970 e consiste em um método computacional para se obter estimativas de parâmetros e realizar testes de hipóteses, não se fazendo muitas suposições sobre a natureza da distribuição dos dados (EFRON e TIBSHIRANI, 1993).

O *bootstrap* é um método de reamostragem baseado na construção de subamostras a partir de uma amostra inicial (Hair Junior et al., 1998). O método é bastante útil quando se deseja avaliar, para um certo estimador, o seu erro padrão, o seu viés, ou ainda quando se quer estimar a distribuição de probabilidade do estimador.

Em estatística, as características retiradas de uma população são obtidas a partir de observações de amostras. Uma vez que a amostra representa a população, suas características deverão fornecer informações sobre as particularidades da população. Segundo Casella e Berger (2010) o *bootstrap* ajuda a aprender sobre as características da amostra (parâmetros estatísticos pela obtenção de reamostras), isto é, é feita a reamostragem a partir da amostra original e utiliza-se essas informações para inferir a respeito de parâmetros da população. O método *bootstrap* pode ser construído de modo paramétrico ou não-paramétrico.

Bootstrap não-paramétrico

No *bootstrap* não-paramétrico, o processo de reamostragem se dá a partir da função de distribuição empírica dos dados, ou dos resíduos, no caso em que há uma estrutura de regressão (GIVENS e HOETING, 2005). Segundo esses autores, cada elemento da amostra tem

a mesma probabilidade $1/n$ de ser selecionado, pois considera-se que a amostra original é independente e identicamente distribuída com uma função de distribuição desconhecida, sendo n o tamanho da amostra.

Para uma amostra $\mathbf{X} = (X_1, X_2, \dots, X_j, \dots, X_n)$ e um estimador $\hat{\theta}(X_1, X_2, \dots, X_n) = \hat{\theta}$, seleciona-se B reamostras, ou amostras *bootstrap*. Em cada uma delas estima-se o parâmetro de interesse pela mesma função usada na amostra original, assim, na i -ésima amostra *bootstrap* tem-se o i -ésimo estimador *bootstrap* dado por $\hat{\theta}_i^*(X_1^*, X_2^*, \dots, X_n^*) = \hat{\theta}_i^*$, em que X_j^* é o j -ésimo elemento da amostra *bootstrap* obtido por um processo de amostragem por reposição da amostra original. As B reamostragens irão fornecer B estimativas, cuja média é denotada por

$$\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*. \quad (1.8)$$

O vício de $\hat{\theta}$ baseado nas B réplicas é então dado por

$$\text{vício} = \bar{\theta}^* - \hat{\theta} \quad (1.9)$$

e o estimador *bootstrap* corrigido por vício é dado por

$$\hat{\theta}_B = 2\hat{\theta} - \bar{\theta}^*. \quad (1.10)$$

A variância do estimador $\hat{\theta}$ é obtida por

$$Var_B^*(\hat{\theta}) = \frac{1}{B-1} \sum_{i=1}^B \left(\hat{\theta}_i^* - \bar{\theta}^* \right)^2. \quad (1.11)$$

A distribuição empírica, contendo B estimativas, é denominada distribuição *bootstrap* do estimador $\hat{\theta}$. Ela pode ser usada, além da obtenção da variância do estimador ou do seu erro padrão, também para a construção de intervalos de confiança e para testes de hipóteses. No caso particular de construção de intervalos de confiança, pelo menos cinco diferentes métodos de estimação por intervalo existem (Efron e Tibshirani, 1993). O mais importante é conhecido por intervalo de confiança com correção acelerada de viés. Nos testes de hipóteses, a hipótese nula é imposta de alguma forma, antes de se obter a distribuição *bootstrap*, que deverá ser usada para a tomada de decisão a respeito da rejeição ou não de H_0 .

O *bootstrap* considerado até agora foi o não-paramétrico, uma vez que não se assume forma para a função densidade de probabilidade ou função distribuição acumulada da população (CASELLA e BERGER, 2010). Por outro lado, existe ainda o método conhecido por *bootstrap* paramétrico, que será detalhado na sequência.

***Bootstrap* paramétrico**

No *bootstrap* paramétrico são feitas suposições a respeito da forma funcional da distribuição da variável aleatória que se está amostrando. Nesse caso, reamostra-se observações da distribuição postulada, usando os valores das estimativas dos parâmetros, obtidas na amostra original, como parâmetros dessa distribuição no processo de geração de pseudo-amostras (DAVISON e HINKLEY, 2008).

Considere que $X_1, X_2, \dots, X_j, \dots, X_n$ seja uma amostra obtida de uma distribuição com função densidade de probabilidade $f(x|\theta)$, em que θ pode ser um vetor de parâmetros. Sendo $\hat{\theta}$ o estimador de máxima verossimilhança (EMV) de θ , obter amostras aleatórias de tamanho n de $f(x|\hat{\theta})$ sendo que

$$X_1^*, X_2^*, \dots, X_j^*, \dots, X_n^* \sim f(x|\hat{\theta}). \quad (1.12)$$

Se forem consideradas B dessas amostras, pode-se estimar a variância de $\hat{\theta}$ utilizando (1.11). Essa distribuição pode ser utilizada para obtenção de intervalos de confiança e com algumas pequenas modificações, também pode ser utilizada para se realizar testes de hipóteses. Essas amostras não são reamostras dos dados, mas amostras aleatórias, obtidas a partir de $f(x|\hat{\theta})$, que, algumas vezes, é chamada de distribuição plugada (CASELLA e BERGER, 2010).

1.1.3 Séries Temporais

Uma série temporal pode ser definida como qualquer conjunto de observações tomadas sequencialmente no tempo. Quando os dados são tomados sequencialmente no tempo há de se esperar que exista uma correlação entre as observações no instante t e em um tempo subsequente $t + h$. Baseado neste fato, alguns modelos usados pelas metodologias estatísticas clássicas não podem ser aplicados, pois, a pressuposição de independência nem sempre é satisfeita. Sendo assim, a melhor abordagem a ser utilizada são as técnicas de séries temporais. Uma classe geral de modelos no domínio do tempo é a dos modelos denominados autoregressivos integrados de médias móveis, ARIMA (p, d, q) , sendo p a ordem da parte autoregressiva do modelo, d o número de diferenças para tornar a série estacionária e q a ordem da parte de médias móveis. Nesse sentido, os estudos de séries temporais são baseados em modelos para séries que visam investigar periodicidades e tendências, bem como fazer previsões.

Modelos de Decomposição

Um modelo clássico de séries temporais supõe que a série temporal Z_t , $t = 1, 2, \dots, N$, possa ser escrita como a soma (modelo aditivo) ou multiplicação (modelo multiplicativo) de suas componentes. As expressões destes modelos, são apresentadas, respectivamente, pelas

equações (1.13) e (1.14),

$$Z_t = T_t + S_t + a_t \quad (1.13)$$

e

$$Z_t = T_t S_t a_t, \quad (1.14)$$

em que $t = 1, 2, \dots, N$; T_t é a componente de tendência, que pode ser entendida como um aumento ou diminuição gradual das observações ao longo do tempo; S_t é a componente sazonal, indicando possíveis flutuações ocorridas em períodos de tempo; a_t é um termo aleatório mostrando as oscilações aleatórias.

Ao se retirar as componentes de tendência e sazonalidade, o que resta é a_t (erro aleatório). Segundo Morettin e Toloi (2004), existindo uma sequência $\{a_t, t \in T\}$ de variáveis independente e indenticamente distribuídas (i.i.d), com média zero e variância constante, tais variáveis são chamadas choques aleatórios e a sequência é denominada ruído branco. O modelo da Eq. (1.13) é chamado aditivo, no qual a sazonalidade varia de maneira independente da componente tendência, ou seja, S_t não depende de T_t . No modelo da Eq. (1.14), o modelo multiplicativo, tais componentes são dependentes e sua variância não é constante. Esse modelo pode ser transformado em um modelo aditivo mediante aplicação de alguma transformação, como por exemplo, o logaritmo.

Modelos de Box e Jenkins

A análise de séries temporais, segundo a metodologia de Box e Jenkins (1976), tem como principal objetivo a realização de previsão. Uma das características fundamentais dessa metodologia é interpretar uma dada série temporal como sendo uma realização de um processo estocástico. Um processo estocástico pode ser definido como o conjunto de todas as possíveis séries temporais de um processo físico observável que se possa ter.

Dentre os diversos métodos e modelos de previsão, os modelos de Box e Jenkins consiste em ajustar modelos ARIMA (p, d, q) , a um conjunto de dados. Esses modelos se caracterizam por serem simples, parcimoniosos e as previsões são bastante precisas. Os modelos se classificam em modelos lineares estacionários e não estacionários.

Alguns operadores são amplamente utilizados no desenvolvimento dos modelos de Box e Jenkins. São eles:

1. B é operador translação para o passado: este operador ocasiona uma defasagem de um período m para trás, cada vez que é utilizado. É definido por:

$$B^m Z_t = Z_{t-m}; \quad (1.15)$$

2. F é o operador translação para o futuro: responsável por uma defasagem de um período de tempo para frente. E definido por:

$$F^m Z_t = Z_{t+m}; \quad (1.16)$$

3. Δ é o operador diferença. Definido por:

$$\Delta Z_t = Z_t - Z_{t-1} = (1 - B)Z_t \Rightarrow \Delta = 1 - B. \quad (1.17)$$

Da Eq. (1.17) segue que:

$$\Delta^n Z_t = (1 - B)^n Z_t \Rightarrow \Delta^n = (1 - B)^n; \quad (1.18)$$

4. S é o operador soma. Definido por:

$$SZ_t = \sum_{j=0}^{\infty} Z_{t-j} = (1 + B + B^2 + \dots)Z_t = (1 - B)^{-1}Z_t \Rightarrow S = \Delta^{-1}; \quad (1.19)$$

Modelos lineares Estacionários

Existem três casos particulares de processos lineares estacionários. São eles: processo autoregressivo de ordem p , representado por $AR(p)$; processo de médias móveis de ordem q , representado por $MA(q)$; processo autoregressivos e de médias móveis de ordens p e q , representados por $ARMA(p, q)$.

1) Modelos Autoregressivos - $AR(p)$

Um modelo autoregressivo de ordem p , $AR(p)$, é escrito em função de seus valores passados e do ruído branco. É denotado por:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t. \quad (1.20)$$

A equação (1.21) pode ser reescrita como:

$$\phi(B)\tilde{Z}_t = a_t \quad (1.21)$$

em que $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ é o polinômio autoregressivo de ordem p .

2) Modelos de Médias Móveis - $MA(q)$

O modelo de médias móveis resulta da combinação linear dos ruídos brancos ocorridos no período corrente e nos períodos passados, podendo ser escrito por

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (1.22)$$

ou

$$\tilde{Z}_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t = \theta(B) a_t, \quad (1.23)$$

sendo $\theta(B)$ o operador de médias móveis estacionário de ordem q .

3) Modelo Autoregressivos e médias móveis - ARMA (p, q)

Ao se combinar o modelo autoregressivo (AR) e o modelo médias móveis (MA), obtém-se o modelo autoregressivo médias móveis ARMA (p, q), sendo p a ordem da parte autoregressiva e q a ordem da parte médias móveis, é escrito da seguinte forma

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (1.24)$$

ou

$$\phi(B) \tilde{Z}_t = \theta(B) a_t. \quad (1.25)$$

sendo $\phi(B)$ e $\theta(B)$, respectivamente, os operadores autoregressivos e de médias móveis.

Modelos Não-Estacionários

A maioria das séries temporais apresenta algum tipo de não estacionariedade. Como a maioria dos procedimentos de séries temporais supõe que estas sejam estacionárias, é necessário transformar os dados originais no caso de não estacionariedade. A transformação mais comumente aplicada consiste em tomar diferenças sucessivas na série original até se obter uma série estacionária. O número de diferenças necessário para tornar uma série estacionária é denominada ordem de integração.

De acordo com Morettin e Tolo (2004), se por meio de um número finito de diferenças, d , a série torna-se estacionária, tem-se o caso de séries não estacionárias homogêneas. Dois são os casos mais comuns de não estacionariedade: séries não estacionárias quanto ao nível, as quais oscilam ao redor de um nível médio durante algum tempo e depois saltam para outro nível temporário.

Neste caso, para torná-las estacionárias uma diferença é suficiente; séries não estacionárias quanto à inclinação, as quais oscilam numa direção por algum tempo e depois mudam,

temporariamente, para outra direção, sendo necessário tomar a segunda diferença. Dentre os modelos propostos para o ajuste deste tipo de dados, temos o modelo autoregressivo integrado médias móveis ARIMA (p, d, q) .

1) Modelo Autoregressivo integrado de médias móveis - ARIMA(p, d, q)

Os modelos na forma ARIMA (p, d, q) são descritos conforme a Eq. (1.26),

$$\phi(B) \Delta^d Z_t = \theta(B) a_t \quad \text{ou} \quad \phi(B)(1 - B)^d Z_t = \theta(B) a_t, \quad (1.26)$$

em que $\phi(B) = 1 - \phi_1(B) - \phi_2(B^2) - \dots - \phi_p(B^p)$ é o operador autoregressivo de ordem p ; Δ^d é o operador diferença simples, d indicando o número de diferenças; $\theta(B) = \theta_1(B) - \theta_2(B^2) - \dots - \theta_q(B^q)$ é o operador de médias móveis de ordem q ; a_t é o resíduo que, eventualmente, pode ser ruído branco.

Funções de autocorrelação (FAC) e Funções de autocorrelação parcial (FACP)

As correlações da FAC indicam a ordem do modelo de médias móveis (MA) e as da FACP indicam a ordem do autoregressivo (AR).

Características da FAC:

- Processo autoregressivo de ordem p - AR(p): constituído por uma mistura de polinômios, exponenciais e senóides amortecidas;
- Processo de médias de ordem q - MA(q): é igual a zero para “lag” maiores do que q ;
- Processo autoregressivo e de médias móveis de ordem p, q - ARMA(p, q): possui (FAC) infinita em extensão, a qual decai de acordo com exponenciais e ou senóides amortecidas após o “lag” $q - p$.

Características da FACP:

- Processo autoregressivo de ordem p -AR (p): possui (FACP) $\phi_{kk} \neq 0$, para $k \leq p$ e $\phi_{kk} = 0$, para $k > p$;
- Processo de médias móveis de ordem q -MA (q): a FACP se comporta de maneira similar à (FAC) de um processo AR (p), ou seja, exponenciais e ou senóides amortecidas;
- Processo autoregressivo e de médias móveis de ordem p, q - ARMA(p, q): a (FACP) se comporta como a (FACP) de um processo MA puro.

Tendência

Supondo-se a ausência de sazonalidade, tem-se o modelo:

$$Z_t = T_t + a_t \quad t = 1, 2, \dots, n \quad (1.27)$$

em que a_t é um ruído branco, com variância σ_a^2 e $n = N/2$. Existem vários métodos para estimar tendência. Os mais utilizados são: ajuste polinomial e suavização (filtro), descritos em Morettin e Tolo (2004). Estimando-se a tendência por meio de $\hat{T}_t$, a série ajustada para tendência ou livre de tendência é dada por:

$$Y_t = Z_t - \hat{T}_t. \quad (1.28)$$

Um procedimento também utilizado para eliminar a tendência consiste em tomar diferenças sucessivas da série original. A primeira diferença é dada por:

$$\Delta Z_t = Z_t - Z_{t-1} \quad (1.29)$$

geralmente é suficiente para deixar a série estacionária.

A construção de gráficos na análise de séries temporais é uma ferramenta importante. Por meio de gráficos é possível identificar características inerentes aos dados, como variabilidades, observações atípicas, sazonalidade e tendência, dentre outras. Mas, como os procedimentos visuais nem sempre são confiáveis, existem testes para confirmar a presença destas componentes.

Os testes para tendência, citados em Morettin e Tolo (2004), são: teste de sequências (Wald-Wolfowitz), teste do sinal (Cox-Stuart) e o teste baseado no coeficiente de correlação de Spearman. Abaixo segue a descrição do teste do sinal de Cox-Stuart, segundo Morettin e Tolo (2004).

Teste do sinal de Cox-Stuart

O teste consiste em dividir a série em dois grupos, nos quais são comparadas as observações (Z_1, Z_{1+c}) , (Z_2, Z_{2+c}) , $\dots$, (Z_{N-c}, Z_N) , em que $c = \frac{N}{2}$ para situações em que N é um número par e $c = \frac{N+1}{2}$, quando N for um número ímpar. É atribuído sinal “+” sempre que $Z_t < Z_{t+c}$ e sinal “-” $Z_t > Z_{t+c}$, ou seja, é a soma de sinais “+” com os sinais “-”. É testada a hipótese bilateral

$$H_0 : P(Z_t < Z_{t+c}) = P(Z_t > Z_{t+c}), \forall i : \text{não existe tendência};$$

$$H_1 : P(Z_t < Z_{t+c}) \neq P(Z_t > Z_{t+c}), \forall i : \text{existe tendência};$$

Definindo T_2 como o número de sinais positivos, rejeita-se H_0 , ou seja, a série apresenta tendência se $T_2 \geq n - t$, em que t é obtido por meio de uma distribuição binomial

com parâmetros ($n = c$, $p = 1/2$) e α um dado nível de significância. Para $n > 20$, utiliza-se a aproximação normal.

Sazonalidade

Da mesma forma que a tendência, a sazonalidade ou periodicidade constitui uma outra forma de não-estacionaridade e deve ser estimada e retirada da série. Sendo $\hat{S}_t$ a estimativa de S_t , a série sazonalmente ajustada considerando um modelo aditivo é

$$Z_t^{SA} = Z_t - \hat{S}_t. \quad (1.30)$$

Priestley (1989) propõe o teste de Fisher, usando para testar a presença de sazonalidade determinística, que é baseado na análise de uma quantidade chamada de periodograma, a qual é dependente das funções seno e cosseno.

O periodograma é uma descrição dos valores observados numa realização de uma série por meio da sobreposição de ondas sinusoidais com várias frequências. A aplicação prática mais óbvia desta decomposição é a de servir de instrumento à identificação de componentes cíclicas ou periódicas.

Conforme Priestley (1989), a função periódica é dada por:

$$I_p(f_i) = \frac{2}{n} \left[\left(\sum_{t=1}^n a_t \cos \frac{2\pi i}{n} t \right) \left(\sum_{t=1}^n a_t \sin \frac{2\pi i}{n} t \right)^2 \right] \quad (1.31)$$

em que $0 < f_i < 1/2$; $t = 1, 2, \dots, n$; $I_p(f_i)$ é a intensidade da frequência f_i . A periodicidade de período $1/f_i$ pode ser observada pela existência de picos na frequência $f_i = 1/n$.

Teste de Fisher

Para se testar as hipóteses de nulidade (H_0), a qual contempla a não existência de sazonalidade, usa-se a estatística

$$g = \frac{\max I_p}{\sum_{p=1}^{N/2} I_p} \quad (1.32)$$

em que I_p é o valor do periodograma, Eq. (1.31), no período p e N é o número de observações da série. A estatística do teste de Fisher, z_α , é dado por:

$$z_\alpha = 1 - \left(\frac{\alpha}{n} \right)^{\frac{1}{n-1}} \quad (1.33)$$

em que $n = N/2$ e α é o nível de significância do teste.

Se $g > z_\alpha$, rejeita-se H_0 , ou seja, a série apresenta periodicidade p .

Teste de Box-Pierce

O teste de Box-Pierce (Box e Pierce, 1970) para verificar se o resíduo é ruído branco é baseado nas k primeiras estimativas das autocorrelações, $\hat{r}_k^2$, dos ruídos. A estatística do teste é dada por:

$$Q(k) = n(n+2) \sum_{j=1}^k \frac{r_k^2}{n-j}. \quad (1.34)$$

Se o modelo ajustado for apropriado $Q(k) \sim \chi_{k-p-q}^2$, em que k é o número de “lags”, p é a ordem da parte autoregressiva do modelo e q é a ordem da parte de médias móveis, a hipótese de ruído branco será aceita se $Q(k) < \chi^2$ tabelado.

Previsões

Uma das formas de utilização de um modelo ajustado é fazer previsões de valores futuros. A previsão de Z_{t+h} , para $h = 1, 2, \dots$ é denotada por $\hat{Z}_t(h)$ e pode ser definida como a esperança condicional de Z_{t+h} , dados todos os valores passados, isto é:

$$\hat{Z}_t(h) = [Z_{t+h}] = E[Z_{t+h} | Z_t, Z_{t-1}, \dots]. \quad (1.35)$$

O erro de previsão é definido por:

$$e_t(h) = Z_{t+h} - \hat{Z}_t(h) \quad (1.36)$$

em que Z_{t+h} é o valor real e $\hat{Z}_t(h)$ é o valor predito.

Escolha do Modelo de Séries Temporais

Um dos métodos que podem ser utilizados para a seleção de modelos é o Critério de Akaike (AIC) e o Critério de Informação Bayesiano (BIC). De acordo com Morettin e Tolo (2004) a identificação do modelo é realizada por meio da série diferenciada. O AIC e o BIC podem ser calculado respectivamente pelas Eq. (1.37) e (1.38):

$$AIC(p, q) = \ln(\hat{\sigma}_{p,q}^2) + 2 \frac{(p+q)}{N} \quad (1.37)$$

$$BIC(p, q) = \ln(\hat{\sigma}_{p,q}^2) + \frac{(p+q)\ln N}{N} \quad (1.38)$$

sendo $\hat{\sigma}_{p,q}^2$ a variância do modelo estimado pelo método de máxima verossimilhança e $p + q$ dado pelo número de parâmetros do modelo estimado (p é a ordem da parte autoregressiva do modelo e q é a ordem da parte de médias móveis).

Referências

- BARIONI, L.G.; ALBERTINI, T.Z.; MEDEIROS, S.R. Modelagem matemática do balanço de gases do efeito estufa na pecuária de corte. **Revista Brasileira de Zootecnia**, Viçosa, v. 40, n. 1, p. 201-212, Set. 2011.
- BOX, G.; PIERCE, D. Distribution of residual auto-correlations in autoregressive-integrated moving average time series models. **Journal American Statistical Association**, Alexandria, v. 65, n. 332, p. 1509-1529, Dec 1970.
- BOX, G.E.P.; JENKINS, G.M. **Time Series Analysis: forecasting and control**. San Francisco: HoldenDay, 1976. 575 p.
- EFRON, B.; TIBSHIRANI, R.J. **An introduction to the bootstrap**. Boca Raton: Chapman & Hall, 1993. 436 p.
- CASELLA, G.; BERGER, R.L. **Inferência estatística**. São Paulo: Cengage Learning, 2010. 588 p.
- CHIB, S.; GREENBERG, E. Understanding the Metropolis-Hastings Algorithm. **The American Statistician**, Alexandria, v. 49, n. 4, p. 327-335, Nov. 1995.
- DAVISON, A.C.; HINKLEY, D.V. **Bootstrap methods and their application**. Cambridge: Cambridge University, 2008. 582 p.
- FEI, W. Evaluation on performances of the bayesian vector autoregression model in regional economic forecasting, **Chinese Journal**, Pequim, v. 10, n. 1, p. 138-143, Dec. 2014.
- FRANCE, J.; GILL, M.; THORNLEY, J.H.M.; ENGLAND, P. A model of nutrient utilization and body composition in beef cattle. **Animal Production**, England, v. 44, p. 371-385, Oct. 1987.
- GAMERMAN, D. **Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. London: Chapman & Hall, 1997, 245 p.
- GELMAN, A.; RUBIN, D.B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v. 7, n. 4, p. 457-511, Nov. 1992.

GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: BERNADO, J.M.; BERGER, J.O.; DAWID, A.P.; SMITH, A.F.M. (Ed.). **Bayesian Statistics 4**. New York: Oxford University Press, 1992. 526 p.

GIVENS, G.H; HOETING, J. A. **Computational statistics**. Hoboken: Wiley-Interscience, 2005. 418 p.

HAARIO, H.; LAINE, M.; MIRA, A.; SAKSMAN, E. DRAM: Efficient Adaptive MCMC. **Statistics and Computing**, New York, v. 16, p. 339-354, 2006.

HAARIO, H; SAKSMAN, E. TAMMINEN, J. An adaptive Metropolis algorithm. **Bernoulli**, Netherlands, v. 2, n. 7, p. 223-242, 2001.

HAIR JUNIOR, J.F.; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E.; TATHAM, R.L. **Análise multivariada de dados**. 5 ed. Porto Alegre: Bookman, 1998. 593 p.

HASTINGS, W.K. Monte Carlo Sampling methods using Markov chains and their applications. **Biometrika**, London, v. 57, n. 1, p. 97-109, 1970.

HEIDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operations Research**, Landing, v. 31, n. 6, p. 1109-1144, 1983.

HOCH, T.; AGABRIEL, J. A mechanistic dynamic model to estimate beef cattle growth and body composition: 1. Model description. **Essex**, United Kingdom, n. 81, p. 1-15, Jun. 2004.

MCPHEE, M.J. OLTJEN, J.W. FADEL, J.G. MAYER, D.G.; SAINZ, R.D. Parameter estimation of fat deposition models in beef steers. **DNA**, New Zelandia, v. 4, p. 115-120, Dec. 2007.

MIRA A. On Metropolis-Hastings algorithms with delayed rejection. **Metron**, Varese, v. 59, n. 3, p. 231-241, 2001.

MORETTIN, P.A.; TOLOI, C.M.C. **Análise de Séries Temporais**. São Paulo: Edgard Blucher, 2004. 535 p.

MORAES, L.E; STRATHE, A.B.; FADEL, J.G.; CASPER, D.P.; KEBREAB, E. Prediction of enteric methane emissions from cattle. **Global Change Biology**, New York, v. 20, n. 1, p. 2140-2148, Nov. 2013.

OLTJEN, J.W.; BYWATER, A.C.; BALDWIN R.L.; GARRETT W.N. Development of a dynamic model of beef cattle growth and composition. **Jornal Animal Science**, Champaign, v. 62, p. 86-97, 1986.

PAULINO, C.D.; TURKMAN, M.A.A.; MURTEIRA, B. **Estatística bayesiana**. Lisboa: Portugal, 2003. 447p.

PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K. CODA: Convergence Diagnosis and Output Analysis for MCMC, **R News**, New Zealand, v. 6, n. 1, p. 7-11, Mar. 2006.

PRIESTLEY, M.B. **Spectral analysis and time series**. London: Academic Press, 1989. 407 p.

RAFTERY, A.L.; LEWIS, S. Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. **Statistical Science**, Hayward, v. 7, n. 4, p. 493-497, Nov. 1992.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014. Disponível em: <<http://www.R-project.org>>. Acesso em: 20 mar. 2014.

SOETAERT, K.; PETZOLDT, T. Inverse Modelling, Sensitivity and Monte Carlo Analysis in R Using Package FME. **Journal of Statistical Software**, Los Angeles, v. 33, n. 3, p. 1-28, Feb. 2010.

STEINFELD, H.; GERBER, P.; WASSENAAR, T.; CASTEL, V.; ROSALES, M.; HAAN, C. **Livestock's long shadow**: Environmental issues and options. Rome, Italy, 2006. 377 p.

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

TEDESCHI, L.O.; SCHOFIELD, P.; PELL, A.N. Determining feed quality for ruminants using in vitro gas production technique. 1. Building an anaerobic fermentation chamber. In **The 4th Workshop on Modeling in Ruminant Nutrition**: Application of the Gas Production Technique, Juiz de Fora, MG, Brazil, 2008, p. 1-16.

UGARTE, M.D.; ADIN, A.; GOICOA, T.; MILITINO, A.F. On fitting spatio-temporal disease mapping models using approximate Bayesian inference, **Statistical Methods in Medical Research**, Bethesda, v. 23, n. 6, p. 507-530, Apr. 2014

ZAPF, A.; BRUNNER, E. KONIETSCHKE, F. A Wild Bootstrap approach for the selection of biomarkers in early diagnostic trials. **BMC Medical Research Methodology**, London, v. 15, n. 43, p. 1-13, May 2015.

WALMSLEY, B.J.; WOLCOTT, M.L.; MCPHEE, M.J. Modeling the relationship between scanned rump and 12th-rib fat in young temperate and tropical bovines: Model development and evaluation. **Jornal Animal Science**, Champaign, v. 88, n. 1, p. 1848-1859, Dec. 2014

WILLIAMS, C.B.; BENNETT, G.L. Application of a computer model to predict optimum slaughter end points for different biological types of feeder cattle. **Jornal Animal Science**, Champaign, v. 73, p. 2903-2915, 1995.

WILLIAMS, C.B.; JENKINS T.G. A dynamic model of metabolizable energy utilization in growing and mature cattle. II. Metabolizable energy utilization for gain. **Jornal Animal Science**, Champaign, v. 81, n. 6, p. 1382-1389, Jun. 2003.

2 COMPARAÇÃO DO AJUSTE DE EQUAÇÕES: UM ESTUDO DE PRODUÇÃO DE GÁS *IN VITRO*

Resumo

Sob algumas condições de produção, a pecuária desempenha um papel primordial na concentração de Gases de Efeito Estufa (GEE) na atmosfera, impondo novos desafios para o setor e para as políticas públicas do país. Especificamente, o gás carbônico e o metano produzido durante a fermentação anaeróbica dos alimentos é uma das principais fontes de GEE em sistemas de produção de ruminantes. A caracterização dos padrões de fermentação ruminal de diferentes alimentos é um passo importante na formulação de dietas balanceadas, que podem reduzir o seu potencial impacto negativo para o meio ambiente. Os objetivos deste estudo foram: *i*) para o conjunto de dados sintético (produção de gases simulada por meio de parâmetros fixos e acréscimo de ruídos representando o erro aleatório) identificar qual tipo de modelagem apresentam maior qualidade por meio de análise frequentista e com um procedimento bayesiano com três equações não-lineares diferentes (equações exponencial, logística e logística bicompartimental) e dois procedimentos de integração: integração numérica de equações não-lineares em forma diferencial (isto é, equações diferenciais ordinárias, EDO) e a forma analítica, integral exata. *ii*) para um conjunto de dados de produção de CO_2 obtidos por meio de fermentação *in vitro* de feno de alfafa, identificar qual tipo de modelagem apresentam maior qualidade por meio de análise frequentista, *bootstrap* e com o procedimento bayesiano com três equações não-lineares diferentes (equações exponencial, logística e logística bicompartimental) e os dois procedimentos de integração: EDO e a forma analítica. Adicionalmente, modelos de séries temporais como ARIMA (auto-regressivo integrado de médias móveis) também foram ajustados para contemplar a dependência residual da produção acumulativa da produção de gases CO_2 . Os principais critérios adotados para avaliar os ajustes dos modelos foram o Erro Quadrático Médio de Predição (EQMP), o Critério de Informação Akaike (AIC) e o Critério de Informação Bayesiano (BIC). Concluiu-se: *i*) para o conjunto de dados sintético, variabilidades de erros aleatórios menores (desvio padrão igual 0.05) as soluções na forma analítica por meio MCMC apresentou comportamento de ajustes melhores e os modelos não-lineares mostraram se adequados para estimar a produção de gases com resíduos normais, homogêneos e independentes. *ii*) para o conjunto de dados de produção de CO_2 obtidos por meio de fermentação *in vitro* de feno de alfafa, a abordagem de séries temporais foi adequada apresentado análise residual com distribuição homogênea, independente e normal (ruído branco). Dentre os modelo de séries temporais o modelo ARIMA (1, 1, 2) teve maior precisão e exatidão com menores valores do EQMP (8, 5⁻⁶), AIC (-7715,9773) e BIC (-13,5665). O modelo ARIMA (1, 1, 2) ajustou a produção acumulativa de medidas de gases CO_2 , atingindo o valor máximo de 1,1066 (mL) no tempo de 47,5 horas e a equação é indicada para estimar a produção de gases.

Palavras-chave: Equação diferencial ordinária; Modelos não-lineares; Pacote FME; Gases de efeito estufa; Monte Carlo via Cadeia de Markov

Abstract

Under some production conditions, livestock play a role in the atmospheric Greenhouse Gases (GHG) concentration, imposing new challenges for the sector and to policy-makers. Specifically, the carbon dioxide and methane produced during the anaerobic fermentation of feeds is a main source of GHG in ruminant production systems. The characterization of patterns

of rumen fermentation of different feeds is an important step in formulating balanced diets that can lower their potential negative impact to the environment. The aim of this study were: *i*) for synthetic dataset (production simulated gases through fixed parameters and noise increase representing the random error) identify which type of modeling have higher quality through frequentist analysis and a procedure Bayesian with three equations different non-linear (exponential equations, logistics and bicompartimental logistics) and two integration procedures: the numerical integration of nonlinear equations in differential form (i.e., ordinary differential equations, EDO) and analytical, integral exact. *ii*) a set of CO_2 production data obtained by fermentation in vitro of alfalfa hay, identify of which type of modeling have higher quality through frequentist analysis, bootstrapping and Bayesian procedure with three equations different non-linear (exponential equations, logistics and bicompartimental logistics) and the two integration procedures: EDO and analytically. In addition, time series models as ARIMA (auto-regressive integrated moving average) have also been adjusted to contemplate the residual dependence of the accumulative production of production of CO_2 . The main criteria used to determine the settings of the models were the mean squared error of prediction (EQMP), the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The conclusion is: *i*) for the set of synthetic data, lower variability of random errors (standard deviation 0.05) the solutions in analytical form through MCMC presented behavior best settings and non-linear models showed adequate to estimate the production of gases with normal, homogeneous and independent residues. *ii*) for all CO_2 production data obtained by in vitro fermentation of alfalfa hay, the time series approach was adequate to represent the residual analysis with homogeneous distribution, independent and normal (white noise). Among the model time series, ARIMA model (1, 1, 2) had greater precision and accuracy with lower values of EQMP (8.5^{-6}), AIC (-7715.9773) and BIC (-13.5665). The ARIMA (1, 1, 2) model showed higher performance in cumulative production setting measures CO_2 gases, reaching the maximum value of 1.1066 (*mL*) in the time of 47,5 hours and the equation is indicated for estimating production GHG.

Keywords: Ordinary differential equations; Nonlinear models; FME package; Greenhouse gases; Markov Chain Monte Carlo

2.1 Introdução

A emissão de gases metano (CH_4), óxido nitroso (N_2O) e gás carbônico (CO_2), potentes Gases de Efeito Estufa (GEE), tem levantado preocupações sobre o aquecimento global da indústria da pecuária. Steinfeld et al. (2006) estimaram que o setor pecuário contribui para cerca de 18% do total das emissões antrópicas de GEE globais. A agência de proteção ambiental dos Estados Unidos [U.S. Environmental Protection Agency (USEPA, 2006)] estima que a contribuição direta da pecuária para as emissões de gases, exceto CO_2 (ou seja, CH_4 e N_2O), respondem por cerca de 7,3 e 7,5% dos valores das emissões globais de GEE para 2010 e 2020, respectivamente. Essas emissões são mais baixas para alguns países industrializados (3,1% do total das emissões de GEE nos Estados Unidos 2009; USEPA, 2011). Para que a indústria cumpra com abordagens de produção ecologicamente corretos são necessárias estimativas confiáveis de emissões de CH_4 , N_2O , CO_2 e balanços de GEE, que por sua vez, depende de dietas que combine

adequadamente as necessidades de nutrientes e de energia dos animais e que sejam nutricionalmente balanceadas.

Modelos baseados em sistemas de equações diferenciais não-lineares podem ajudar a fornecer um entendimento quantitativo da complexidade de fenômenos físicos ou biológicos (Calderhead and Girolami, 2011; Tedeschi et al., 2013). Modelos matemáticos nutricionais são representações abstratas da realidade que são úteis para fazer previsões testáveis com diferentes níveis de detalhes e são importantes ferramentas de balanceamento de dietas e podem estimar a emissão de GEE sobre diferentes condições de produção (Tedeschi, 2006). Na presença de dependência entre as observações, ordenamente no tempo, modelos não-lineares podem não representar todas as variações dos dados como os de acúmulos de GEE. Análise de séries temporais podem representar melhor o comportamento dos dados e obter as previsões do volume produzido de gases acumulativos produzido nas fermentações *in vitro* ao longo de períodos subsequentes, usando métodos de Box e Jenkins, como o modelo ARIMA (auto-regressivo integrado médias móveis) e suas variações. A escolha do melhor modelo e estimativas confiáveis dos parâmetros são necessários.

Métodos baseados em Monte Carlo via Cadeia de Markov (MCMC) estão se tornando mais populares do que os métodos tradicionais para obter o intervalo das estimativas. As vantagens do MCMC em comparação com os métodos tradicionais incluem estimativas pontuais com maior probabilidade de ocorrência e intervalos de credibilidades para os parâmetros dos modelos. Destaca-se ainda, pontos fortes ao lidar com a não-linearidade, interdependência dos parâmetros em modelos complexos e incorporar o conhecimento *a priori* no que diz respeito aos parâmetros por meio de experiências anteriores (Soetaert e Petzoldt, 2010; Wang e Chen, 2013).

Técnicas de reamostragem *bootstrap* são muito utilizadas para representar os intervalos de confiança dos parâmetros e também o estudo da estrutura de correlação entre eles. As reamostragens permitem diferentes alternativas para se encontrar desvios padrões e a distribuição assintótica dos parâmetros de interesse.

Há falta do uso de previsões da produção de gases sobre o aspecto de séries temporais, além de metodologias *bootstrap* e MCMC para as análises de dados no desenvolvimento e aplicações de modelos de nutrição animal com modelos não-lineares. Os objetivos deste estudo foram: *i*) para o conjunto de dados sintético (produção de gases simulada por meio de parâmetros fixos e acréscimo de ruídos representando o erro aleatório) identificar qual tipo de modelagem apresentam maior qualidade por meio de análise frequentista e com um procedimento bayesiano [MCMC com a incorporação da Rejeição por Atraso e o Metropolis Adaptativo (RAMA)] com três equações não-lineares diferentes (equações exponencial, logística e logística bicompartimental) e dois procedimentos de integração: integração numérica de equações não-lineares em forma diferencial (isto é, equações diferenciais ordinárias, EDO) e a forma analítica, integral exata; *ii*) para um conjunto de dados de produção de CO_2 obtidos por meio de fermentação *in vitro* de feno de alfafa, identificar qual tipo de modelagem apresentam maior qualidade por meio de análise frequentista, *bootstrap* e com o procedimento bayesiano com as três equações não-lineares di-

ferentes (equações exponencial, logístico e logístico bicompartimental) e os dois procedimentos de integração: EDO e a forma analítica. Adicionalmente, modelos de séries temporais como ARIMA (auto-regressivo integrado de médias móveis) também foram ajustados, para contemplar a dependência residual da produção acumulativa da produção de gases CO_2 .

2.2 Material e Métodos

Na seção 2.2.1 foi discutido o uso de modelos não-lineares na sua forma analítica e na seção 2.2.2 o uso de sistema de equações diferenciais ordinárias (EDO). As análises foram realizadas utilizando um conjunto de dados sintético e um conjunto de dados contendo medições *in vitro* de pressão de CO_2 (seção 2.2.3). Os parâmetros para ambas as formas, analítica e ODE, foram estimadas usando análise frequentista com o método de Levenberg-Marquardt (seção 2.2.4) e MCMC (seção 2.2.6) implementado nas funções *modFit* e *modMCMC*, respectivamente, do pacote FME. Adicionalmente para o conjunto de dados de gases CO_2 reais foram realizadas análises *bootstrap* (seção 2.2.5) e uma modelagem por meio de séries temporais (seção 2.2.10). A análise de sensibilidade e linearidade de medição dos parâmetros (seção 2.2.7) foram realizados e uma descrição sobre avaliação dos modelos é apresentada na seção 2.2.8. Um diagrama em forma de caixa é apresentado na Figura 2.1 para representar os passos do desenvolvimento dessa metodologia.

Figura 2.1 – Diagrama com a composição metodológica da proposta desse estudo

2.2.1 Equação não-linear na forma analítica (integral exata)

As equações (2.1) e (2.2) mostram a forma básica das funções não-lineares exponenciais e da logística com uma defasagem, respectivamente. Nestas equações, o parâmetro θ_1 representa a assíntota, volume total de gases (mL); o parâmetro θ_2 representa a taxa de fermentação fracionada (h^{-1}); e o parâmetro θ_3 representa a fase de latência - o *lag time* (h , Schofield et al., 1994).

Em termos nutricionais, os carboidratos podem ser classificados como fibrosos (CF) e não-fibrosos (CNF). Os primeiros, representados pela celulose e hemicelulose, juntamente com a lignina, apresentam disponibilidade nutricional incompleta e variável; são lentamente digeridos e ocupam espaço no trato gastrointestinal (TGI). Os CNF, representados pelos açúcares solúveis, amido e pectina, apresentam disponibilidade nutricional praticamente completa (aproximadamente 98%) e constante entre os alimentos e são rapidamente digeridos no TGI (Van Soest, 1967; Fernandes et al.; 2003).

O “*Cornell Net Carbohydrate and Protein System*” (CNCPS) classifica os carboidratos totais em quatro frações de acordo com as taxas de degradação no rúmen: fração A corresponde aos carboidratos solúveis em água, representados principalmente pela glicose e sacarose, que são rapidamente fermentados pelas bactérias utilizadoras de CNF. A fração B_1 é composta por amido e pectina, rapidamente fermentados pelas bactérias utilizadoras de CNF. As frações B_2 e C correspondem às frações potencialmente degradável e não degradável da fibra insolúvel em detergente neutro (FDN, Sniffen et al., 1992; NRC, 1996).

A equação (2.3) mostra a função não-linear logística bicompartimental com um único tempo de *lag* com base nas derivações de Schofield et al. (1994). A interpretação biológica é que θ_1 reflete o volume de gases produzido pela degradação da fração $A + B_1$ do CNCPS; θ_2 a taxa de degradação da fração $A + B_1$ (CNF); θ_3 é a fase de latência; θ_4 é o volume de gases produzidos pela degradação da fração B_2 do sistema de Cornell (CF); e θ_5 é a taxa de degradação da fração B_2 .

$$y(t) = \begin{cases} \theta_1(1 - e^{-\theta_2(t-\theta_3)}), & \forall t \geq \theta_3 \\ 0, & \forall t < \theta_3 \end{cases} \quad (2.1)$$

$$y(t) = \begin{cases} \frac{\theta_1}{1+e^{(2+4\theta_2(\theta_3-t))}}, & \forall t \geq \theta_3 \\ 0, & \forall t < \theta_3 \end{cases} \quad (2.2)$$

$$y(t) = \begin{cases} \frac{\theta_1}{1+e^{(2+4\theta_2(\theta_3-t))}} + \frac{\theta_4}{1+e^{(2+4\theta_5(\theta_3-t))}}, & \forall t \geq \theta_3 \\ 0, & \forall t < \theta_3 \end{cases} \quad (2.3)$$

As principais vantagens do modelo de regressão não-linear é que frequentemente existe interpretação para a maioria de seus parâmetros. Esses parâmetros então passam ser o foco da investigação que, na sua forma mais simples, consiste em determinar intervalos de confiança e testar hipóteses. No entanto, o interesse é a necessidade de fazer inferência sobre uma função dos parâmetros. Embora a forma analítica das equações não-lineares permitam obter intervalos de credibilidade e conduzir testes de hipóteses é possível fazer inferências utilizando a sua forma derivada, com ODE partindo de um valor inicial.

2.2.2 Equação diferencial ordinária

Equações (2.4), (2.5) e (2.6) tem a forma EDO para a exponencial, logística e logística bicompartimental, respectivamente.

$$\frac{\partial y}{\partial t} = \begin{cases} \theta_2(\theta_1 - y), & \forall t \geq \theta_3 \\ 0, & \forall t < \theta_3, \end{cases} \quad (2.4)$$

em que y é o valor da variável dependente no valor atual do tempo t .

$$\frac{\partial y}{\partial t} = \begin{cases} \frac{4\theta_1\theta_2e^{4\theta_2(\theta_3-t)+2}}{(e^{4\theta_2(\theta_3-t)+2}+1)^2}, & \forall t \geq \theta_3 \\ 0, & \forall t < \theta_3 \end{cases} \quad (2.5)$$

$$\frac{\partial y}{\partial t} = \begin{cases} \frac{4\theta_4\theta_5e^{4\theta_5(\theta_3-t)+2}}{(e^{4\theta_5(\theta_3-t)+2}+1)^2} + \frac{4\theta_1\theta_2e^{4\theta_2(\theta_3-t)+2}}{(e^{4\theta_2(\theta_3-t)+2}+1)^2}, & \forall t \geq \theta_3 \\ 0, & \forall t < \theta_3 \end{cases} \quad (2.6)$$

Dedução da equação exponencial:

$$\frac{\partial y}{\partial t} = \begin{cases} \theta_2(\theta_1 - y), & \forall t \geq \theta_3 \\ 0, & \forall t < \theta_3, \end{cases} \quad (2.7)$$

$$\Rightarrow \frac{\partial y}{\partial t} = \theta_1\theta_2(1 - \frac{y}{\theta_1}), \forall t \geq \theta_3 \quad (2.8)$$

$$\Rightarrow \int \frac{\partial y}{-\frac{y}{\theta_1} + 1} = \int_{\theta_3}^t \theta_1\theta_2 dt, \forall t \geq \theta_3 \quad (2.9)$$

Usando a regra de integração,

$$\int \frac{\partial y}{(ay + b)} = \frac{1}{a} \ln(ay + b) \text{ e ajustando } a = -1/\theta_1 \text{ e } b = 1, \quad (2.10)$$

tem-se que:

$$\Rightarrow -\theta_1 \ln(-\frac{y}{\theta_1} + 1) = \theta_1\theta_2(t - \theta_3), \forall t \geq \theta_3 \quad (2.11)$$

$$\Rightarrow \ln\left(-\frac{y}{\theta_1} + 1\right) = -\theta_2(t - \theta_3), \forall t \geq \theta_3 \quad (2.12)$$

$$\Rightarrow -\frac{y}{\theta_1} + 1 = e^{\{-\theta_2(t-\theta_3)\}}, \forall t \geq \theta_3 \quad (2.13)$$

$$\Rightarrow \frac{y}{\theta_1} = 1 - e^{\{-\theta_2(t-\theta_3)\}}, \forall t \geq \theta_3 \quad (2.14)$$

$$\Rightarrow y = \begin{cases} \theta_1(1 - e^{-\theta_2(t-\theta_3)}), & \forall t \geq \theta_3 \\ 0, & \forall t < \theta_3, \end{cases} \quad (2.15)$$

em que a Eq. 2.15 corresponde a Eq. 2.1 equivalente da forma analítica, conforme demonstrado.

As equações (2.5) e (2.6) podem ser obtidas de forma similar ou usando a função *deriv* do pacote *stats* (R Core Team, 2014). A diferença no ajuste de uma equação diferencial [Eq. (2.4), (2.5) e (2.6)] para uma equação analítica [Eq. (2.1), (2.2) e (2.3)], respectivamente, é a informação do valor inicial que é indicado nos modelos de equações diferenciais. Os modelos foram implementados no *software* R (Development Core Team, versão 3.1, 2014). Para calcular a solução numérica de problemas de valores iniciais com sistemas de EDO de primeira ordem foi utilizado a função *ode()* do pacote *deSolve* (Soetaert et al., 2010). A solução foi realizada pelo método de integração *lsoda* com erro de tolerância absoluto e relativo de 10^{-6} .

2.2.3 Banco de dados

A análise estatística foi realizada utilizando dois conjuntos de dados. O primeiro conjunto de dados foi composto por dados sintético e o segundo de dados contido em medições *in vitro* da produção de gás de CO_2 de fermentação anaeróbica de feno de alfafa.

Banco de dados sintético. Neste cenário, foram simulados dados usando funções não-lineares [Eqs. (2.1), (2.2), (2.3), (2.4), (2.5) e (2.6), respectivamente] e os seus parâmetros foram mantidos constantes (isto é, fixo) e os erros aleatórios foram adicionados utilizando uma distribuição normal, $N(\mu, \sigma)$, com $\mu = 0$ e $\sigma = 0,05$ (para baixa variação) e $\sigma = 0,2$ (em alta variação). A Tabela 2.1 resume os valores de cada parâmetro do modelo simulado.

O valor inicial do modelo no momento t_0 foi fixada no valor zero, isto é, $f(t_0) = 0$, para evitar quaisquer respostas negativas, uma vez que a produção de gás na fermentação do rúmen nunca toma valores negativos. Então, para todos os modelos e ruídos considerados, $f(t) > 0, \forall t$.

Tabela 2.1 – Simulação do banco de dados sintético para comparar inferências estatísticas usando a forma analítica e EDO com solução frequentista e MCMC

Modelo	y_0 (caso EDO)	Designação dos valores do parâmetros					Erro
		θ_1	θ_2	θ_3	θ_4	θ_5	
exponencial	0	2	0,0700	0,0500	-	-	N(0, 0,05)
logístico	0	2	0,0700	0,0500	-	-	
logístico bicompartimental	0	10,6500	0,1400	0,4300	10	0,0370	N(0, 0,2)
exponencial	0	2	0,0700	0,5000	-	-	
logístico	0	2	0,0700	0,5000	-	-	
logístico bicompartimental	0	10,6500	0,1400	0,4300	10	0,0370	

Conjunto de dados baseados na produção de gás CO_2 de ruminantes *in vitro*. As fermentações anaeróbicas *in vitro* foram obtidas com os compostos de feno de alfafa de alta qualidade produzidas em fazendas do sudeste do Texas (TX), fornecida pelo setor de equideocultura da Texas A & M University, com a finalidade de medir a pressão gerada pela produção de CO_2 por fermentação microbiana ruminal. O inóculo ruminal utilizado foi composto por uma mistura dos fluídos ruminais de dois novilhos da raça Britânico-cruzado, mantidos em pasto de forragens mistas com qualidade média recebendo uma vez ao dia concentrado comercial para bovinos em crescimento. A produção acumulada de CO_2 foi medida de 5 em 5 minutos durante o período de 48 h de incubação, que representou a digestão da matéria orgânica no rúmen e forneceu uma indicação confiável da dinâmica de fermentação que podem ser utilizados para avaliar a produção de CH_4 . Os dados foram cedidos pelo *Department of Animal Science* - Texas A & M University (EUA). Mais detalhes sobre a metodologia *in vitro* foi discutido por Tedeschi et al. (2008, 2009).

2.2.4 Processos numéricos iterativos: algoritmos de otimização

O algoritmo de Levenberg-Marquardt propõe uma solução entre o algoritmo de gradientes e o método iterativo de Gauss-Newton que é baseada no método dos mínimos quadrados (MORE, 1978). Seja o modelo não-linear F definido de forma geral para o vetor de parâmetros $\theta = [\theta_1 \ \theta_2 \ \dots \ \theta_j]'$ e para o vetor de variáveis regressoras da i -ésima unidade amostral $Z_i' = [Z_{1i} \ Z_{2i} \ \dots \ Z_{pi}]$ por

$$Y_i = F_i(\theta, Z_i) + \epsilon_i. \quad (2.16)$$

Em notação matricial o modelo pode ser representado por:

$$Y = F(\theta) + \epsilon. \quad (2.17)$$

O método Levenberg-Marquardt baseia-se na minimização da soma de quadrados dos resíduos $L(\theta) = \epsilon'\epsilon$. Substituindo $\epsilon = Y - F$ e derivando com relação a θ , tem-se

$$L(\theta) = \epsilon'\epsilon = (Y - F)'(Y - F) = Y'Y - 2Y'F + F'F, \quad (2.18)$$

$$\frac{\partial L}{\partial \theta} = \frac{-\partial 2Y'F}{\partial \theta} + \frac{F'F}{\partial \theta} \quad (2.19)$$

em que expressou-se o vetor do modelo $F(\theta)$, simplesmente por F . Mas,

$$\frac{-\partial 2Y'F}{\partial \theta} = \frac{-\partial 2Y'F}{\partial F} \cdot \frac{F}{\partial \theta} = -2Y'X \quad (2.20)$$

em que $X = F/\partial \theta$ é a matriz de derivadas parciais, em que cada coluna é formada pela derivada linear em relação aos parâmetros. Por outro lado,

$$\frac{-\partial F'F}{\partial \theta} = \frac{\partial F'F}{\partial F} \cdot \frac{F}{\partial \theta} = -2F'X. \quad (2.21)$$

Logo,

$$\frac{\partial L}{\partial \theta} = -2Y'X + 2F'X. \quad (2.22)$$

Igualando a zero a primeira derivada, tem-se a equação normal para o modelos não-lineares:

$$X'F = X'Y. \quad (2.23)$$

Como F e X são funções de θ , então não existe uma forma fechada para a solução da Eq. (2.23). Em geral usa-se um processo iterativo, no qual informa-se um valor inicial para o vetor de parâmetros que é melhorado continuamente até que a soma de quadrados de resíduos $e'e$ seja minimizada. Para o valor corrente (k -ésimo passo do processo iterativo) do vetor de parâmetros, deve-se calcular a matriz X e estimar o vetor de resíduos por $e = Y - F(\theta_k)$. No ponto inicial ($k = 0$) é avaliado a matriz X e o vetor de resíduos, considerando o valor arbitrário do vetor de parâmetros especificado. Se $SQE(\theta_k) = e'e$ for a soma de quadrados dos resíduos avaliada na k -ésima iteração, então X e Y são usados para calcular um vetor Δ de tal forma que

$$SQR(\theta_k + \lambda \Delta) < SQE(\theta_k), \quad (2.24)$$

para uma constante λ qualquer. Os métodos de solução da Eq. (2.24) diferem na forma de como Δ é calculada para proporcionar as trocas no vetor de parâmetros. Um resumo de alguns critérios são descritos abaixo:

$$\begin{cases} \text{Gauss-Newton} & \Delta = (X'X)^{-1}X'e \\ \text{Gradiente} & \Delta = X'e \\ \text{Levenberg-Marquardt} & \Delta = [X'X + \delta \text{diag}(X'X)]^{-1}X'e, \end{cases} \quad (2.25)$$

em que $(X'X)^{-1}$ é uma inversa generalizada.

O método de Levenberg-Marquardt mantém um compromisso entre o método de Gauss-Newton e o método Gradiente, se $\delta \rightarrow 0$, existe uma aproximação ao método de Gauss-Newton e se $\delta \rightarrow \infty$, existe uma aproximação ao método do Gradiente. Por padrão, a função *modFit* do pacote *FME* (Soetaert et al., 2010) inicia δ com valores de acordo com a precisão da máquina (64 bits, $\delta = 10^{-7}$). Em termos práticos, as estimativas desses métodos são bem próximas, no entanto, a eficiência de qualquer um deles depende do valor inicial atribuído. Dessa forma, a escolha de um bom valor inicial torna-se mais importante que o método iterativo utilizado para a convergência, Mazucheli e Achcar (2002).

2.2.5 Bootstrap não-paramétrico

Foram realizadas reamostragem de *bootstrap* não-paramétrico no conjunto de dados de gases *in vitro* CO_2 e calculado o intervalo de confiança *Biased Corrected Percentile Bootstrap* (BCPB) descrito no referencial teórico, Capítulo 01. As distribuições de probabilidades dos parâmetros foram representadas graficamente.

Reamostragens com as funções analíticas: Os valores dos parâmetros para as funções analíticas foram estimadas usando o algoritmo de otimização disponível na função *gls* do pacote *nlme*, Pinheiro et al. (2014), com 2.000 reamostragens usando o algoritmo de otimização de Gauss-Newton. Os critérios de parada foram controlados nos argumentos da função *gls* tomando o número de iterações máxima igual a 500 e o erro absoluto igual 10^{-6} (padrão do R).

Reamostragens com as EDO: Para as reamostragens com as EDO, o cálculo da solução numérica de problemas de valores iniciais de primeira ordem foi utilizado a função *ode()* do pacote *deSolve* (Soetaert et al., 2010). A solução foi realizada pelo método de integração *lsoda* com erro de tolerância absoluto e relativo de 10^{-6} . Para estimação dos parâmetros usou a função *modFit* do pacote *FME* (Soetaert et al., 2010), com 2.000 reamostragem usando o método de otimização Levenberg-Marquardt. Os critérios de parada foram controlados nos argumentos da função *modFit* tomando o número de iterações máxima igual a 1.200 e o erro absoluto igual 10^{-6} .

Ao final das estimação das reamostras foram calculados o intervalo de confiança *Biased Corrected Percentile Bootstrap* (BCPB), obtido conforme descrito no referencial teórico. As médias dos vetores com as reamostragens de *bootstrap* foram usadas para avaliar o modelo.

2.2.6 Simulação Monte Carlo via Cadeia de Markov

Métodos bayesianos foram usados para obter a distribuição de probabilidade dos parâmetros para ambas as formas de análises, EDO e analítica, usando as funções não-lineares exponencial, logística e logística bicompartimental. A função *modMCMC* implementa o método de MCMC que usa o procedimento RAMA para aumentar o número de aceites das execuções e atualizações das informações transcorridas durante o aceite de novos valores dos parâmetros e da correlação entre eles (Haario et al., 2006; Laine, 2008). O método de amostragem MCMC baseado numa distribuição conhecida por construção de uma cadeia de Markov tem a distribuição

desejada conforme a sua distribuição de equilíbrio. Assim, em vez de uma estimativa de um vetor de parâmetros, obtém-se um conjunto de vetores que representam a distribuição assintótica desses parâmetros (Soetaert e Petzoldt, 2010).

A função *modMCMC* tem demonstrado ser uma ferramenta de modelação inversa robusta para a resolução de MCMC usando o *software* R (Wu e Liu, 2012). As mensurações $\mathbf{Y}$ foram definidas como indicado abaixo:

$$\mathbf{Y} = f(\mathbf{t}, \boldsymbol{\theta}) + \boldsymbol{\epsilon}, \quad (2.26)$$

$$\boldsymbol{\epsilon} \sim N(0, \sigma^2) \quad (2.27)$$

em que $f(\mathbf{t}, \boldsymbol{\theta})$ é uma equação não-linear, $\mathbf{t}$ é a variável independente (tempo), $\boldsymbol{\theta}$ o vetor de parâmetros e $\boldsymbol{\epsilon}$ é aditivo, independente com distribuição normal, com média zero e variância σ^2 desconhecida.

No primeiro momento, para o conjunto de dados sintético, a distribuição à *priori* para $\boldsymbol{\theta}$, $p_{pri}(\boldsymbol{\theta})$ foi assumida como não informativa, baseado em Martínez-Teruel et al. (2009). Enquanto que para a inversa da variância (σ^{-2}) a distribuição Gama foi usada como uma distribuição *priori* conforme a equação (2.28):

$$p_{pri}(\sigma^{-2}) \sim \Gamma\left(\frac{n_0}{2}, \frac{n_0}{2} S_0^2\right), \quad (2.28)$$

sendo $n_0 = wvar.n$, em que *wvar* é uma ponderação inicial da variância do modelo; n o número de observações (tamanho da amostra) e S_0^2 é a variância inicial do modelo (como retornado no resumo da função *modFit* pelo método de Levenberg-Marquardt). Então a *posteriori* completa foi estimada como:

$$p(\boldsymbol{\theta}, \sigma^2, \mathbf{Y}) \propto \exp\left\{-\frac{1}{2} \left(\frac{SS(\boldsymbol{\theta})}{\sigma^2}\right)\right\} p_{pri}(\boldsymbol{\theta}) \propto (\sigma^{-2})^{\frac{n+n_0}{2}-1} \exp\left\{-\frac{1}{2} [(SS(\boldsymbol{\theta}) + n_0 S_0^2)]\right\} \sigma^{-2} \quad (2.29)$$

em que SS é a função da soma de quadrados $SS(\boldsymbol{\theta}) = \sum_{i=1}^n (y_i - f(\mathbf{t}, \boldsymbol{\theta})_i)^2$; $i = 1, 2, \dots, n$, sendo n o tamanho da amostra.

A cada passo do MCMC, a inversa da variância do erro (*posteriori* condicional completa de σ^{-2}) é amostrado por meio de uma distribuição Gama, como indicada na Eq. (2.30), Gelman et al. (2004):

$$p(\sigma^{-2} | (\mathbf{Y}, \boldsymbol{\theta})) \sim \Gamma\left(\frac{n_0 + n}{2}, \frac{n_0 S_0^2 + SS(\boldsymbol{\theta})}{2}\right), \quad (2.30)$$

e a *posteriori* completa condicional de $\boldsymbol{\theta}$ é calculada como

$$p(\boldsymbol{\theta} | \mathbf{Y}, \sigma^2) \propto \exp\left\{-\frac{1}{2} \left(\frac{SS(\boldsymbol{\theta})}{\sigma^2}\right)\right\}. \quad (2.31)$$

No segundo momento, para o conjunto de dados com mensurações em fermentações anaeróbica *in vitro*, assumiu-se para os parâmetros, θ , *priori* com distribuição Gausiana, conforme resultados obtidos por meio da metodologia *bootstrap* realizada para o mesmo conjunto de dados. Para a inversa da variância do erro (σ^{-2}) uma distribuição Gama foi usada como na Eq. (2.28). A *posteriori* completa foi descrita como Eq. (2.32):

$$p(\theta, \sigma^2 | \mathbf{Y}, \mathbf{t}) \propto (\sigma^{-2})^{\frac{n+n_0}{2}-1} \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n \frac{(y_i - f(\mathbf{t}, \theta)_i)^2}{\sigma^2} + \frac{n_0 S_0^2}{2} \sigma^{-2} + \sum_{j=1}^k \frac{(\theta_j - \mu_{\theta_j})^2}{\sigma_{\theta_j}^2} \right] \right\}, \quad (2.32)$$

em que $i = 1, 2, 3, \dots, n$, sendo n o tamanho da amostra; $j = 1, 2, \dots, k$, k representa o número de parâmetros aleatórios do modelo; μ_{θ_j} e $\sigma_{\theta_j}^2$ representam a média e a variância do parâmetro θ_j ; σ^2 é a variância do modelo; $n_0 = wvar.n$, em que *wvar* é a ponderação da variância inicial e S_0^2 é a variância inicial. Então, assumindo um valor fixo para σ^2 , a *posteriori* condicional completa de θ tem a forma como mostra a Eq. (2.33)

$$p(\theta | \sigma^2, \mathbf{Y}, \mathbf{t}) \propto \exp \left\{ -\frac{1}{2} \left[\sum_{i=1}^n \frac{(y_i - f(\mathbf{t}, \theta)_i)^2}{\sigma^2} + \sum_{j=1}^k \frac{(\theta_j - \mu_{\theta_j})^2}{\sigma_{\theta_j}^2} \right] \right\}, \quad (2.33)$$

A *posteriori* condicional completa de σ^{-2} tem a mesma forma apresentada na Eq. (2.30). Definindo

$$SS_{pri}(\theta) = \sum_{j=1}^k \frac{(\theta_j - \mu_{\theta_j})^2}{\sigma_{\theta_j}^2} \quad (2.34)$$

pode-se escrever a Eq. (2.33) como

$$p(\theta | \sigma^2, \mathbf{Y}, \mathbf{t}) \propto \exp \left\{ -\frac{1}{2} \left[\frac{SS(\theta)}{\sigma^2} + SS_{pri}(\theta) \right] \right\}, \quad (2.35)$$

e a taxa *posteriori* necessaria na probabilidade de aceitação do Metropolis-Hastings foi calculada como

$$\frac{p(\theta^* | \mathbf{Y}, \sigma^2)}{p(\theta | \mathbf{Y}, \sigma^2)} = \exp \left\{ -\frac{1}{2} \left[\frac{SS(\theta^*)}{\sigma^2} - \frac{SS(\theta)}{\sigma^2} \right] + \frac{1}{2} [SS_{pri}(\theta^*) - SS_{pri}(\theta)] \right\}. \quad (2.36)$$

A combinação dos métodos, RA e MA (RAMA), implementado por Soetaert e Petzoldt (2010) do pacote FME foi usada para atualizar a matriz de covariância nesse presente estudo e usada para gerar valores de parâmetros candidatos. No método MA, a geração de novos valores candidato dos parâmetros foi feita de modo mais eficiente por meio do ajuste da distribuição sugerida para o tamanho e forma da distribuição alvo. Isto foi realizado por meio da geração de novos parâmetros com uma matriz de covariância proposta que foi estimada pelos parâmetros gerados e salvos na cadeia até aquele passo (Soetaert e Petzoldt, 2010). Durante o método RA, os novos valores dos parâmetros foram gerados a partir do último valor aceito pelo dimensionamento da matriz de covariância. A probabilidade de aceitação deste novo conjunto depende dos

candidatos até agora propostos e rejeitados, de maneira que a reversibilidade da cadeia de Markov foi preservada. Isso fornece um diagnóstico sistemático quando o processo de adaptação tem um início lento (Haario et al., 2006).

Para a estrutura de matriz de covariância inicial foi considerada a aproximação numérica Jacobiana ($\mathbf{J}$), então, com base na teoria linear, o parâmetro de covariância foi estimado como $(\mathbf{J}^T \mathbf{J})^{-1} S^2$, em que S^2 é a soma de quadrado residual do melhor ajuste. Com o melhor ajuste, $(\mathbf{J}^T \mathbf{J}) \approx 0,5 \mathbf{H}$, em que $\mathbf{H}$ é a matriz Hessiana (Soetaert e Petzoldt, 2010; Press et al. 1992).

Algoritmo MA:

1. Iniciou-se com um vetor inicial $\boldsymbol{\theta}^0$ e uma estrutura de matriz de covariância proposta $\mathbf{C}^{(1)} = \mathbf{C}_0$ para a primeira fase, obtidos pelo ajuste clássico pela função *modFit* usando o método Levenberg-Marquardt. Selecionou-se um fator de escala $covscale = 2,4^2/k$ em que k representa dimensão do número de parâmetros dos modelos, conforme proposto por Germann et al. (2004). Esses autores mostraram que esta escolha de escala otimiza as propriedades de mistura da busca do Metropolis-Hasting no caso de alvos de distribuições Gaussianas. Para a regularização da covariância um número $\varepsilon = 10^{-16}$ foi adicionado na diagonal da matriz de covariância para evitar que ela se tornasse singular. O valor do período de não-adaptação inicial foi tomado como $burninglength = 7.000$. Escolheu a priori dos parâmetros para $\boldsymbol{\theta}$ e os parâmetros para a variância para dos erros (S_0^2 e $wvar0$).
2. Etapa Metropolis-Hastings: Dado o atual valor $\boldsymbol{\theta}$, gerou um novo valor $\boldsymbol{\theta}^*$ de uma distribuição Gaussiana multivariada $N(\boldsymbol{\theta}, \mathbf{C})$, calculou a Soma de quadrados $SS(\boldsymbol{\theta}^*)$ e calculou a soma de quadrados da priori, $SS_{pri}(\boldsymbol{\theta}^*)$. O novo valor foi aceito se

$$u \leq \min \left\{ 1, \exp \left\{ -\frac{1}{2} \left[\frac{SS(\boldsymbol{\theta}^*)}{\sigma^2} - \frac{SS(\boldsymbol{\theta})}{\sigma^2} \right] + \frac{1}{2} [SS_{pri}(\boldsymbol{\theta}^*) - SS_{pri}(\boldsymbol{\theta})] \right\} \right\} \quad (2.37)$$

em que u foi gerado de uma distribuição Uniforme $U(0, 1)$ e σ^2 é o atual valor da variância do erro.

3. Atualizou σ^2 com a obtenção direta da distribuição condicional de $p(\sigma^2 | \boldsymbol{\theta}, y)$, isto é, obtido por $1/\sigma^2$ da distribuição Gama

$$1/\sigma^2 \sim \Gamma \left(\frac{n_0 + n}{2}, \frac{n_0 S_0^2 + SS(\boldsymbol{\theta})}{2} \right) \quad (2.38)$$

4. Após o período inicial de simulação, dito para $i \geq burninglength$ adaptou-se a matriz de covariância $\mathbf{C} = cov(\boldsymbol{\theta}^0, \dots, \boldsymbol{\theta}^i) covscale + \mathbf{I} \varepsilon$.
5. Retornou-se para o passo (2) até que valores suficientes terem sido amostrados.

A combinação dos dois métodos foi realizado de modo que na primeira fase do RA foi adaptado da mesma maneira que no AM: a covariância C_n^1 foi computada por meio dos pontos da cadeia amostrados, não importando em que fase estes pontos foram aceitos na trajetória da amostragem. A covariância C_n^i do segundo estágio ($i = 2$) foi calculado com uma nova escala, usando o padrão do *software R*. Para o segundo estágio de aceitação foi calculado

$$\frac{q_1(\theta^{**}, \theta^*)}{q_1(\theta, \theta^*)} = \exp \left\{ -\frac{1}{2}(\theta^{**} - \theta^*)'(C^{(1)})^{-1}(\theta^{**} - \theta^*) + \frac{1}{2}(\theta - \theta^*)'(C^{(1)})^{-1}(\theta - \theta^*) \right\}. \quad (2.39)$$

No presente trabalho, a monitoração da convergência das cadeias foi realizada por meio de análise gráfica, critério de Geweke (1992) e o critério de Raftery e Lewis (1992). Os testes de diagnósticos de convergência e saídas de análises do pacote *coda* (Plummer et al., 2006) foram usados formalmente para avaliar a convergência da cadeia. O cálculo dos Erros Padrão de Monte Carlo (EPMC) foi realizada pelo pacote *mcmcse* (Flegal e Hughes, 2012). Foi usado um total de 1.007.000 iterações (*niter*), dos quais 7.000 iterações iniciais foram descartados (*burninglength*). Das 1.000.000 iterações restantes apenas 5.000 iterações foram retidas na saída da cadeia (*outputlength*), ou seja, com 200 passos entre um valor para outro.

Com o conjunto de dados sintético, foi utilizada uma cadeia de Markov de cada (forma integral exata e na forma EDO). Para o conjunto de dados com mensurações de fermentação anaeróbica *in vitro*, duas cadeias de Markov de cada (forma integral exata e na forma EDO) foram usadas. A Tabela 2.2 apresenta os valores das distribuições *a priori* tomadas para as mensurações de gases *in vitro*. Utilizou-se o pacote FME do software R (R Development Core Team, 2014), versão 3.1.0, para as construções das cadeias MCMC.

Tabela 2.2 – Informações *prioris* usadas e *posteriores* geradas pelos modelos não-lineares por meio da metodologia bayesiana - produção de gás CO_2 de ruminantes *in vitro*

Modelo	parâmetros	Priori	
		EDO	Analítica
exponencial	θ_1	$\sim N(1,1096, 0, 0007^2)$	$\sim N(1,1096, 0, 0007^2)$
	θ_2	$\sim N(0,0955, 0, 0004^2)$	$\sim N(0,0955, 0, 0004^2)$
	θ_3	$\sim N(-3,9482, 0, 0372^2)$	$\sim N(-3,9499, 0, 0372^2)$
	σ^2	$\sim \Gamma(288,5, 0,0210)$	$\sim \Gamma(288,5, 0,0201)$
logístico	θ_1	$\sim N(1,0878, 0, 0013^2)$	$\sim N(1,0878, 0, 0013^2)$
	θ_2	$\sim N(0,0384, 0, 0003^2)$	$\sim N(0,0384, 0, 0003^2)$
	θ_3	$\sim N(-9,8228, 0, 1566^2)$	$\sim N(-9,8228, 0, 1566^2)$
	σ^2	$\sim \Gamma(288,5, 0,0300)$	$\sim \Gamma(288,5, 0,0300)$
logístico bicompartimental	θ_1	$\sim N(0,8078, 0, 0050^2)$	$\sim N(0,7830, 0, 0042^2)$
	θ_2	$\sim N(0,0657, 0, 0002^2)$	$\sim N(0,0646, 0, 0008^2)$
	θ_3	$\sim N(-5,9937, 0, 5000^2)$	$\sim N(-5,8035, 0, 0952^2)$
	θ_4	$\sim N(0,3528, 0, 0022^2)$	$\sim N(0,3395, 0, 0038^2)$
	θ_5	$\sim N(0,0249, 0, 0002^2)$	$\sim N(0,0246, 0, 0002^2)$
	σ^2	$\sim \Gamma(288,5, 0,0050)$	$\sim \Gamma(288,5, 0,0050)$

2.2.7 Análise de sensibilidade e Colinearidade

A função *sensFun* do pacote FME foi usado para estimar sensibilidade local das EDO para os valores dos parâmetros. Quando a função de sensibilidade é aplicada em conjunto

com os dados observados, para cada ponto, as derivadas dos correspondentes valores modelados com os respectivos valores dos parâmetros são calculados.

As sensibilidades dos parâmetros do modelo são medidas adimensionais representados por uma matriz sensibilidade $S_{i,j}$ calculada por

$$S_{i,j} = \frac{\partial y_i}{\partial \theta_j} \frac{\Delta \theta_j}{\Delta y_i}, \quad (2.40)$$

em que y_i é a i -ésima variável de estado do modelo (em um determinado instante de tempo); θ_j é o j -ésimo parâmetro do modelo; Δy_i é a escala da variável y_i ; $\Delta \theta_j$ é a escala do parâmetro θ_j . Quanto maior for o valor da sensibilidade absoluta, mais importante é o parâmetro para o modelo, assim, as magnitudes dos valores calculados pela função de sensibilidade podem ser utilizadas para classificar a importância dos parâmetros sobre as variáveis de estado.

A colinearidade ou identificabilidade para todas as combinações de parâmetros foram estimados pela função de *collin* do pacote FME, de acordo com Brun et al. (2001). O cálculo é baseado tomando previamente a função de sensibilidade, Eq. (2.40), como argumento. Para qualquer subconjunto de colunas da matriz de sensibilidade, a colinearidade, γ , foi definida por Brun et al. (2001) como:

$$\gamma = \frac{1}{\sqrt{\min(EV(\hat{\mathbf{S}}^T \hat{\mathbf{S}}))}}, \quad (2.41)$$

sendo

$$\hat{\mathbf{S}}_{ij} = \frac{S_{ij}}{\sqrt{\sum_{j=1}^p S_{ij}^2}}, \quad (2.42)$$

em que $\hat{\mathbf{S}}$ contém as colunas da matriz de sensibilidade que correspondem aos parâmetros incluídos no conjunto e EV é a função que calcula os autovalores. O índice de colinearidade é igual a 1 se as colunas são ortogonais, e o conjunto é identificável. O índice tende ao infinito se colunas da matriz de sensibilidade são linearmente dependentes.

2.2.8 Avaliação de modelos

Nessa seção foram apresentadas várias técnicas para comparação matemática dos modelos. Essas técnicas são necessárias para o apoio as decisões, dar sustentabilidade para demonstrar o acerto de um modelo, apresentar evidências para promover a sua aceitação e utilização para determinados propósitos.

Todas as previsões obtidas por meio do ajustes dos modelos podem ser reescritas em uma forma de regressão linear simples, representada pelos valores observados e preditos

conforme a Eq. (2.43):

$$Y_i = \beta_0 + \beta_1 f(X_1, \dots, X_p)_i + \varepsilon_i, \quad (2.43)$$

em que $f(X_1, \dots, X_p)_i$ é a variável dependente (preditora) representada pelas i -ésima predição dos modelo e X_p a p -ésima variável de entrada do modelo; Y_i é o i -ésimo valores observado; β_0 e β_1 são o intercepto e a inclinação da reta, respectivamente; ε_i é o i -ésimo erro aleatório que é independente e identicamente distribuído $\sim N(0, \sigma^2)$. A regressão foi avaliada de acordo com a seguinte hipótese estatística:

$$H_0 : \beta_0 = 0 \text{ e } H_0 : \beta_1 = 0 \text{ e } H_a : \text{ não } H_0. \quad (2.44)$$

Se as hipóteses de nulidade não forem rejeitadas, pode-se concluir que as equações estimaram com precisão os valores observados. A inclinação e o intercepto foram avaliadas separadamente para identificar os problemas específicos de ajuste. Para todas as comparações, o valor de 0,05 foi estabelecido como o nível crítico de probabilidade para o erro tipo I . Os parâmetros β_0 e β_1 foram avaliadas separadamente e simultaneamente, para avaliar se o viés foi representado por uma constante (avaliou-se pela diferença do intercepto do valor paramétrico zero) ou por uma tendência de percentagem (avaliou-se o desvio de inclinação do valor paramétrico 1). O estimador de máxima verossimilhança da variância residual σ^2 foi calculado conforme a Eq. (2.45),

$$\text{EQMP} = \frac{\sum_{i=1}^n (Y_i - f(X_1, \dots, X_p)_i)^2}{n}, \quad (2.45)$$

em que $f(X_1, \dots, X_p)_i$ é a variável dependente (preditora); Y_i são os valores observados; e n o número de observações. O EQMP representa o erro quadrado médio de predição, menores valores para essa estimativa indica o melhor modelo na avaliação.

A estimativa de precisão foi determinada pelo Coeficiente de Correlação de Concordância (CCC), ou índice de reprodutibilidade, conforme descrito em Tedeschi (2006) e representada pela Eq. (2.46):

$$\hat{\rho}_c = \hat{\rho} C_b, \quad (2.46)$$

em que o primeiro componente ($\hat{\rho}$) é a estimativa do coeficiente de correlação de *Pearson* dado pela Eq. (2.47):

$$\rho = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}, \quad (2.47)$$

sendo Y_i os valores observados e $\hat{Y}_i$ são os valores preditos pelo modelo de regressão linear. O segundo componente (C_b) é o fator de correção do viés que indica quão longe a linha de regressão

desvia-se da unicidade da inclinação (45°), tal como indicado nas Eqs. (2.48) e (2.49):

$$C_b = \frac{2}{\left(\nu + \frac{1}{\nu} + \mu^2\right)}, \quad (2.48)$$

$$\nu = \frac{s_Y}{s_{f(X_1, \dots, X_p)}} \text{ e } \mu = \frac{\bar{Y} - \bar{f}(X_1, \dots, X_p)}{\sqrt{s_Y s_{f(X_1, \dots, X_p)}}}, \quad (2.49)$$

em que $\bar{Y}$ e $\bar{f}(X_1, \dots, X_p)$ são as médias dos valores observados e preditos, respectivamente; s_Y e $s_{f(X_1, \dots, X_p)}$ são os desvios padrão de Y e $f(X_1, \dots, X_p)$, respectivamente. O fator C_b varia de 0 à 1, quando o C_b é igual a 1 indica que não tem desvio do ajuste da reta de regressão linear ($Y = f(X_1, \dots, X_p)$). O CCC indica modelos com boa exatidão e precisão (quando perto de 1,0) ou modelos com problemas de reprodutibilidade (quando perto de 0,0).

Outros métodos que podem ser utilizados para a seleção de modelos é o Critério de Akaike (AIC), o Critério de Akaike corrigido AIC_c e o Critério de Informação Bayesiano (BIC). O AIC, AIC_c e o BIC foram calculados respectivamente pelas Eq. (2.50), (2.51) e (2.52):

$$AIC = \ln(\hat{\sigma}^2) + 2 \frac{(npar)}{n} \quad (2.50)$$

$$AIC_c = n \cdot \ln \left(\frac{SQE}{n} \right) + 2(npar + 1) + \frac{2(npar + 1)(npar + 2)}{n - (npar + 2)} \quad (2.51)$$

em que a última da expressão da Eq. (2.51) é o termo de correção, sugerido para pequenas amostras, e o

$$BIC = \ln(\hat{\sigma}^2) + \frac{(npar) \ln(n)}{n}, \quad (2.52)$$

sendo $\hat{\sigma}^2$ a variância do modelo estimada; $npar$ o número de parâmetros do modelo estimado e n o número de observações; SQE a soma de quadrados residual. O melhor desempenho do modelo são fornecidos pelos menores valores de AIC, AIC_c e BIC.

Para a avaliação dos modelos também foi usada o fator de eficiência da modelagem (MEF), calculado como:

$$MEF = 1 - \frac{\sum_{i=1}^n (Y_i - f(X_1, \dots, X_p)_i)^2}{\sum_{i=1}^n n(Y_i - \bar{Y})^2}. \quad (2.53)$$

O MEF é interpretado como a proporção da variação explicada pela reta $Y = f(X_1, \dots, X_p)$. Para um perfeito ajuste, as estatísticas Eq. 2.47 e 2.53 resulta no valor igual a 1,0.

O coeficiente de determinação (CD) do modelo é a razão entre a variância total dos dados observados para o quadrado da diferença entre o modelo predito e média dos dados

observados, que foi calculado usando a seguinte estrutura:

$$CD = 1 - \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sum (f(X_1, \dots, X_p)_i - \bar{Y})^2}. \quad (2.54)$$

A estatística CD explica a proporção da variância total dos valores observados explicados pelos dados previsto (Tedeschi, 2006; Loague e Green, 1991). Quanto mais próximo de 1,00 melhor são as predições do modelo.

2.2.9 Teste de aleatorização de Mantel

O teste de aleatorização de Mantel foi usado para comparar duas matrizes de distâncias ou similaridades, Mantel (1967). Para o presente estudo, o interesse foi comparar as matrizes de distância entre as correlações dos parâmetros obtidos pelos ajustes da metodologia *bootstrap* e MCMC. Para descrição do teste, suponha duas matrizes simétricas, $\mathbf{M}$ e $\mathbf{E}$, de distâncias compostas por quatro objetos cada uma delas. As matrizes $\mathbf{M}$ e $\mathbf{E}$ de dimensões 4×4 , na qual a entrada na i -ésima linha e j -ésima coluna é uma medida da distância entre o objeto i e o objeto j de cada conjunto foram descritas conforme (2.55):

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} \text{ e } \mathbf{E} = \begin{bmatrix} e_{11} & e_{12} & e_{13} & e_{14} \\ e_{21} & e_{22} & e_{23} & e_{24} \\ e_{31} & e_{32} & e_{33} & e_{34} \\ e_{41} & e_{42} & e_{43} & e_{44} \end{bmatrix}. \quad (2.55)$$

O teste de Mantel avalia se os elementos em $\mathbf{M}$ e $\mathbf{E}$ mostram alguma correlação significativa e o teste estatístico é calculado usando a correlação de *Pearson*, r_{ME} , entre os elementos correspondentes, combinando m_{11} com e_{11} , m_{12} com e_{12} , $\dots$, m_{44} com e_{44} . A estatística do teste é comparada com a distribuição de Z que é obtida tomando os objetos (linhas ou colunas) em uma ordem aleatória para uma das matrizes, razão pela qual ele é chamado de um teste de aleatorização (Manly, 2008). Para o teste de aleatorização, a matriz $\mathbf{M}$, correspondente a matriz de distância entre os objetos da matriz de correlação calculada por meio de *bootstrap* não-paramétrico nesse estudo, foi mantida fixa. E ordens aleatórias foram escolhidas para os objetos da matriz $\mathbf{E}$ representadas pelas distâncias entre as correlações dos parâmetros obtidos por meio da metodologia MCMC. A correlação r_{ME} tem a interpretação usual em termos do relacionamento entre as duas medidas de distâncias, variando de -1 a +1, com $r = -1$ indicando uma correlação negativa perfeita, $r = 0$ indicando nenhuma correlação e $r = +1$ indicando uma correlação positiva perfeita (Manly, 2008). Para o presente estudo, as distâncias entre os objetos da matriz de correlação entre os parâmetros ajustados pela metodologia *bootstrap* (D_{boot}) e por MCMC (D_{mcmc}) foram tomados de acordo com as Eqs. (2.56) e (2.57), respectivamente:

$$D_{boot} = [\mathbf{1}]_{p \times p} - [\mathbf{R}_{boot}]_{p \times p} \quad (2.56)$$

e

$$D_{mcmc} = [\mathbf{1}]_{p \times p} - [\mathbf{R}_{mcmc}]_{p \times p}, \quad (2.57)$$

em que $[\mathbf{1}]_{p \times p}$ é uma matriz quadrada de 1's com dimensão $p \times p$, sendo p o número de parâmetros do modelo; $[R_{boot}]_{p \times p}$ e $[R_{mcmc}]_{p \times p}$ as matrizes de correlações ajustadas pelas metodologias *bootstrap* e MCMC, respectivamente. Para realização do teste de aleatorização de Mantel foi usado o pacote *biotools* (Silva, 2015), pela função *mantelTest* com um total de 1.000 permutações. A hipótese testada foi:

$$H_0 : \rho = 0 \times H_a : \rho > 0. \quad (2.58)$$

2.2.10 Análise de Séries Temporais

Nessa seção, foi descrito os procedimentos para o ajuste de modelos univariados com séries temporais. A série utilizada nesse estudo foi:

- série de produção de gases CO_2 , coletadas por um período de tempo entre 0 até 47,42 horas, com variação de 0,0833 h entre uma observação e outra, totalizando 570 observações. As observações referentes ao tempo 47,5 à 48,00 horas foi reservada para previsões;

A metodologia aplicada no ajuste dos modelos obedeceu os seguintes passos:

- 1- construção do gráfico da série original para captar indícios de tendência, sazonalidade, variações atípicas e dados discrepantes;
- 2- construção dos gráficos das funções de autocorrelação (FAC) e autocorrelação parcial (FACP), também para observar características de tendências e (ou) sazonalidade;
- 3- aplicação dos testes do sinal (Cox-Stuart) para verificar existência de tendência e de Fisher para verificar sazonalidade;
- 4- caso, na etapa (3), seja detectada tendência e (ou) sazonalidade, foi aplicada diferenças para eliminar estas componentes. Normalmente, para eliminar tendências, 1 ou 2 diferenças são necessárias e para sazonalidade a diferença é da ordem do período;
- 5- nesta etapa, constroem-se os gráficos das funções de autocorrelação (fac) e autocorrelação parcial (fACP) da série diferenciada para estimar as ordens do modelo;
- 6- ajuste dos modelos ARIMA pelo método da máxima verossimilhança e verificação da validade das estimativas;
- 7- aplicação do teste de Box-Pierce para verificar se os resíduos são ruído branco;

- 8- construção das previsões;
- 9- aplicação do critério de AIC, BIC e do EQMP para escolha do melhor modelo ajustado aos dados. Escolhe-se o modelo que apresentar os menores valores de AIC, BIC e EQMP;
- 10- equação do modelo ajustado com as respectivas estimativas;

O programa utilizado para realizar as análises de séries temporais foi o STATISTICA[®] (2001).

2.3 Resultados e Discussões

Dividiu-se essa seção em duas etapas: uma referente aos resultados do banco de dados sintéticos e outra ao banco de dados com as mensurações de CO_2 *in vitro* a partir de métodos de incubação com líquido ruminal.

2.3.1 Banco de dados sintéticos

Nesse presente trabalho, a monitoração da convergência das cadeias foi realizada por meio de métodos formais (testes pelo critério de Geweke e do critério de Raftery e Lewis) e informal (análise gráfica - apresentada no Anexo A). Para o critério de Geweke, existirão evidências contra a convergência da cadeia, se o valor p estimado for menor que o nível de significância pré-fixado pelo pesquisador, enquanto que, no critério de Raftery e Lewis, caracterizam como não convergência da cadeia se o fator de dependência (FD) apresentar valor maior que “cinco”. Verificou que o critério de Raftery e Lewis e critério de Geweke apresentaram convergências dos parâmetros dos modelos ($FD < 5$ e ao nível de 5% de significância, respectivamente), exceto para forma EDO na equação exponencial para o parâmetro θ_3 , Tabela 2.3. O critério de Raftery e Lewis também detectou não convergência para o caso da forma EDO para o modelo logístico, quando aplicado um erro aleatório maior, $N(0; 0, 2^2)$. Baixos valores do erro padrão Monte Carlo (EPMC) é verificado para todos os parâmetros dos modelos testados, exceto na forma EDO na equação exponencial para o parâmetro θ_3 (Tabelas 2.4 e 2.5). A atribuição de distribuições Uniformes como *priori*, com intervalos mais estreitos podem melhorar a convergência dessas equações que não conseguiram atingir convergência, usando métodos MCMC pelo pacote FME.

Para a forma analítica, as estimativas dos parâmetros foram relativamente próximas para θ_1 e θ_2 tanto no caso MCMC como na análise frequentista da equação não-linear exponencial, ou seja, as maiores porcentagens de mudança em relação ao verdadeiro valor foram de $-4,95\%$ (θ_1) e $24,57\%$ (θ_2 , Tabela 2.4). O parâmetro θ_3 , que representa a fase de latência foi subestimado pelas análises frequentista e MCMC (equação exponencial). As estimativas dos parâmetros para o modelo logístico bicompartimental foram similares tanto pelo método frequentista quanto pelo método MCMC. No caso da metodologia MCMC o intervalo de credibilidade conseguiu contemplar o verdadeiro valor de θ_1 , que representa o valor assintótico da curva, para todas as equações não-lineares.

Tabela 2.3 – Fator de dependência do critério de Raftery e Lewis (FD) e valor p do critério de Geweke dos modelos não-lineares - dados sintético

Modelo	Erro	Parâmetro	Forma			
			Analítica		EDO	
			FD	Geweke (valor p)	FD	Geweke (valor p)
exponencial	$N(0, 0, 05^2)$	θ_1	1,0200	0,1228	1,0300	0,1734
		θ_2	1,0200	0,1662	1,0300	0,3332
		θ_3	1,0200	0,9282	69,4000	0,0000
		σ^2	1,0000	0,9847	1,0300	0,8063
logístico	$N(0, 0, 05^2)$	θ_1	1,0500	0,9526	0,9820	0,8103
		θ_2	1,0200	0,2133	0,9990	0,8720
		θ_3	1,0300	0,4524	1,0300	0,9582
		σ^2	1,0000	0,2669	0,9990	0,5000
logístico bicompartimental	$N(0, 0, 05^2)$	θ_1	0,9990	0,7479	1,0100	0,7114
		θ_2	0,9820	0,4293	1,0200	0,9218
		θ_3	1,0200	0,3571	1,0200	0,4774
		θ_4	0,9660	0,6912	1,0200	0,8327
		θ_5	0,9590	0,8908	1,0200	0,5200
		σ^2	0,9820	0,3392	0,9990	0,9085
exponencial	$N(0, 0, 2^2)$	θ_1	1,0200	0,9561	1,0000	0,1011
		θ_2	1,0700	0,6019	1,0789	0,0860
		θ_3	1,1000	0,3042	42,1052	0,0000
		σ^2	1,0300	0,5200	1,0000	0,3628
logístico	$N(0, 0, 2^2)$	θ_1	0,9823	0,6667	1,0789	0,2198
		θ_2	0,9986	0,8551	16,1000	0,6956
		θ_3	1,0320	0,6754	14,4000	0,3851
		σ^2	0,9986	0,9403	1,1842	0,9128
logístico bicompartimental	$N(0, 0, 2^2)$	θ_1	1,0526	0,2905	1,0000	0,1560
		θ_2	1,0526	0,3192	1,1842	0,2651
		θ_3	1,0526	0,6156	1,1842	0,4396
		θ_4	1,0000	0,3714	1,0526	0,1426
		θ_5	1,0526	0,0808	1,0526	0,3480
		σ^2	1,1842	0,2532	1,0000	0,3589

Tabela 2.4 – Resultados da comparação dos modelos não-lineares por análise frequentista e MCMC na forma analítica - *priori* não informativa - banco de dados sintético

Modelo		Frequentista			MCMC					Sobreposição	
		erro	Par.	Intervalo de confiança	Média MCMC	Intervalo de credibilidade	% aceitação	[tempo (h)]			
Média Freq.	(inferior; superior)			(inferior; superior)		EPMC					
exponencial $\theta_1=2,00$ $\theta_2=0,07$ $\theta_3=0,05$	$N(0; 0, 05^2)$	θ_1	1,9646	(1,9414; 1,9878)	1,9661	(1,9206; 2,0150)	0,0018	83,6%	[0,59 h]	48,80%	
		θ_2	0,0758	(0,0723; 0,0793)	0,0757	(0,0688; 0,0833)	0,0003			48,27%	
		θ_3	-0,6678	(-0,9206; -0,4150)	-0,6906	(-1,2394; -0,2009)	0,0258			48,68%	
	$N(0; 0, 2^2)$	θ_1	1,9010	(1,8261; 1,9759)	1,9138	(1,7664; 2,1217)	0,0122	81,42%	[0,76 h]	41,60%	
		θ_2	0,0870	(0,0718; 0,1021)	0,0872	(0,0546; 0,1253)	0,0026			42,96%	
		θ_3	-0,6978	(-1,6206; 0,2250)	-0,9411	(-3,6569; 0,9655)	0,1719			39,92%	
	logístico $\theta_1=2,00$ $\theta_2=0,07$ $\theta_3=0,05$	$N(0; 0, 05^2)$	θ_1	1,9929	(1,9793; 2,0005)	1,9847	(1,9672; 2,0122)	0,0007	73,11%	[1,10 h]	56,44%
			θ_2	0,0766	(0,0728; 0,0804)	0,0815	(0,0774; 0,0863)	0,0001			22,22%
			θ_3	-0,3917	(-0,7555; -0,0279)	0,0904	(0,0009; 0,3671)	0,0097			0,00%
$N(0; 0, 2^2)$		θ_1	1,9760	(1,9348; 2,0172)	1,9821	(1,9023; 2,0653)	0,0036	80,44%	[1,00 h]	49,63%	
		θ_2	0,0827	(0,0682; 0,0972)	0,0855	(0,0616; 0,11429)	0,0011			55,03%	
		θ_3	0,0000	(-1,1980; 1,1980)	-0,3743	(-2,7986; 1,3001)	0,0810			58,45%	
logístico bicompartimental $\theta_1=10,62$ $\theta_2=0,14$ $\theta_3=0,43$ $\theta_4=10,00$ $\theta_5=0,037$		$N(0; 0, 05^2)$	θ_1	10,6539	(10,5754; 10,7324)	10,6537	(10,4929; 10,8141)	0,0011	81,00%	[0,50 h]	48,87%
			θ_2	0,1397	(0,1376; 0,1418)	0,1397	(0,1353; 0,1440)	0,0000			48,27%
			θ_3	-0,5814	(-0,6260; -0,5348)	-0,5813	(-0,6764; -0,4884)	0,0006			99,60%
	θ_4		9,9539	(9,8814; 10,0264)	9,9544	(9,8059; 10,1001)	0,0010	49,25%			
	θ_5		0,0370	(0,0368; 0,0372)	0,0370	(0,0364; 0,0370)	0,0000	25,00%			
	$N(0; 0, 2^2)$	θ_1	10,7530	(10,4330; 11,0730)	10,7465	(10,1034; 11,3852)	0,0044	80,88%	[0,55 h]	49,92%	
		θ_2	0,1388	(0,1304; 0,1472)	0,1392	(0,1231; 0,1563)	0,0001			50,6%	
		θ_3	-0,6149	(-0,8020; -0,4278)	-0,6206	(-1,0062; -0,2674)	0,0027			73,88%	
		θ_4	9,8177	(9,5217; 10,1137)	9,8240	(9,2369; 10,4198)	0,0041			50,04%	
		θ_5	0,0372	(0,0361; 0,0383)	0,0372	(0,0350; 0,0395)	0,0000			48,88%	

Os resultados não atingiram convergência pelas duas metodologias (Frequentista e MCMC) no modelo exponencial (Tabela 2.5) nos ajustes das EDO's. No caso frequentista as estimativas das matrizes de covariância foram singulares. No caso da função logística bicompartimental, mesmo usando uma *priori* não informativa boas estimativas para o período de latência, θ_3 , foi encontrada no ajuste por MCMC (Tabela 2.5). A monitoração informal da convergência da cadeia de cada parâmetro dos modelos se realizou por meio de técnicas gráficas baseadas nos traços das cadeias. Os traços das cadeias e a distribuição de probabilidade dos parâmetros são mostrados no apêndice A, Figura 4.5 à Figura 4.28. As análises gráficas confirmam os resultados encontrados pelos critérios de Geweke e pelo critério de Raftery e Lewis.

Martínez-Teruel et al. (2009) estimaram curvas de crescimento exponencial de diferente subprodutos (limão e vários tipos de palha de cevada tratada e não tratada) que foram degradadas por incubação *in vitro* com o fluido do rúmen extraído a partir de dois rebanhos de cabras *Murciano-Granadina*, no estado de Murcia na Espanha (ESP). Um rebanho foi alimentado com feno de alfafa e o outro alimentado com palha de cevada. As curvas de degradação usando função não-linear exponencial foram ajustadas com a análise frequentista e comparadas com os resultados de uma análise bayesiana. Martínez-Teruel et al. (2009) também usou uma distribuição *a priori* não informativa como informação dos parâmetros para o cálculo da distribuição *posteriori*. Os resultados mostraram que a utilização da análise bayesiana suavizaram as estimativas dos ajustes frequentista usando mínimos quadrados. Os autores concluíram que os ajustes frequentistas não processam adequadamente as restrições envolvidas e levam a resultados biologicamente inaceitáveis.

Tabela 2.5 – Resultados da comparação dos modelos não-lineares por análise frequentista e MCMC na forma EDO - *priori* não informativa - banco de dados sintético

Modelo (Valores Sintéticos)	erro	Par.	Frequentista		MCMC		% aceitação	[tempo (h)]	Sobreposição
			Média Freq.	Intervalo de confiança (inferior; superior)	Média MCMC	Intervalo de credibilidade (inferior; superior)			
exponencial $\theta_1=2,00$ $\theta_2=0,07$ $\theta_3=0,05$	$N(0; 0, 05^2)$	θ_1	*** ¹	*** ¹	1,9375	(1,9339; 2,0183)	0,0004	[9,46 h]	-
		θ_2	*** ¹	*** ¹	0,0830	(0,0686; 0,0775)	0,0000		-
		θ_3	*** ¹	*** ¹	-9×10^{12}	$(-\infty; \infty)$	5×10^{17}		-
	$N(0; 0, 2^2)$	θ_1	*** ¹	*** ¹	1,7519	(1,7707; 2,0574)	0,0001	[8,96 h]	-
		θ_2	*** ¹	*** ¹	1,8761	(0,0659; 0,1095)	0,0002		-
		θ_3	*** ¹	*** ¹	-2×10^{16}	$(-\infty; \infty)$	5×10^{17}		-
logística $\theta_1=2,00$ $\theta_2=0,07$ $\theta_3=0,05$	$N(0; 0, 05^2)$	θ_1	1,9996	(1,9549; 2,0443)	2,2118	(1,9739; 2,1013)	0,0008	[7,26 h]	48,80%
		θ_2	0,0716	(0,0695; 0,0737)	0,0707	(0,0639; 0,0774)	0,0000		30,87%
		θ_3	0,0176	(-1,0168; 1,0520)	0,0235	(-1,2933; 0,6327)	0,0136		16,37%
	$N(0; 0, 2^2)$	θ_1	1,9838	(1,7781; 2,1895)	2,7178	(1,8950; 3,5193)	0,2318	[8,1 h]	25,32%
		θ_2	0,0779	(0,0576; 0,0982)	0,0722	(0,0411; 0,1096)	0,0013		59,27%
		θ_3	0,1203	(0,1025; 0,1566)	-1,8358	(-1,9118; 2,2418)	0,3814		0,39%
logístico bicompartimental $\theta_1=10,62$ $\theta_2=0,14$ $\theta_3=0,43$ $\theta_4=10,00$ $\theta_5=0,037$	$N(0; 0, 05^2)$	θ_1	10,6600	(10,4948; 10,8252)	10,6456	(10,4771; 10,8164)	0,0022	[11,46 h]	91,60%
		θ_2	0,1330	(0,1289; 0,1371)	0,1401	(0,1360; 0,1441)	0,0000		7,23%
		θ_3	0,2295	(0,2153; 0,3566)	0,4274	(0,3399; 0,5061)	0,0008		5,74%
		θ_4	9,9970	(9,9965; 9,9975)	9,9593	(9,7990; 10,1170)	0,0020		0,31%
		θ_5	0,0368	(0,0364; 0,0372)	0,0370	(0,0365; 0,0376)	0,0000		68,10%
	$N(0; 0, 2^2)$	θ_1	10,6500	(10,6483; 10,6517)	10,7146	(10,0879; 11,3993)	0,0086	[11,70]	0,26%
		θ_2	0,1364	(0,1306; 0,1421)	0,1404	(0,1241; 0,1561)	0,0002		36,05%
		θ_3	0,2870	(0,0215; 0,5525)	0,4096	(0,0468; 0,7078)	0,0044		73,68%
		θ_4	9,9060	(9,6496; 10,1624)	9,8457	(9,2523; 10,4321)	0,0078		43,46%
		θ_5	0,0373	(0,0364; 0,0395)	0,0373	(0,0350; 0,0396)	0,0000		37,58%

Na tabela 2.6 foi apresentada uma análise comparativa para avaliação dos ajustes dos modelos. Valores de AIC e BIC foram semelhantes usando as duas metodologias para a equação exponencial. Para o modelo logístico, a análise frequentista apresentou melhor desempenho. O modelo logístico bicompartimental teve menores valores de AIC e BIC quando ajustado pelo método MCMC em todos os cenários de estudos.

Tabela 2.6 – Avaliação de modelos não-lineares metodologias frequentista e MCMC (*priori* não informativa) - dados sintéticos

Forma	Metodologia	Modelo	Erro aleatório	EQMP	AIC	BIC
Analítica	Frequentista	exponencial	$N(0; 0, 05^2)$	0,0000	-135,5731	-129,8976
		logístico		0,0002	-119,0599	-113,3844
		logístico bicompartimental		0,0006	-140,5175	-134,8421
	MCMC	exponencial	$N(0; 0, 05^2)$	0,0000	-135,5718	-129,8963
		logístico		0,0459	-116,1531	-110,4776
		logístico bicompartimental		0,0004	-140,6337	-134,9582
	Frequentista	exponencial	$N(0; 0, 2^2)$	0,0000	-3,8166	1,8588
		logístico		0,0066	-1,4325	4,2428
		logístico bicompartimental		0,0000	-4,7723	0,9030
	MCMC	exponencial	$N(0; 0, 2^2)$	0,0190	-3,8165	-1,8589
		logístico		0,0968	-0,8663	4,8090
		logístico bicompartimental		0,0000	-4,7733	0,9020
EDO	Frequentista	exponencial	$N(0; 0, 05^2)$	***	***	***
		logístico		0,0002	-139,8491	-134,1737
		logístico bicompartimental		0,1083	52,4231	58,0985
	MCMC	exponencial	$N(0; 0, 05^2)$	***	***	***
		logístico		0,0032	-139,6933	-134,0178
		logístico bicompartimental		0,1020	36,1576	41,8331
	Frequentista	exponencial	$N(0; 0, 2^2)$	***	***	***
		logístico		0,0000	-3,9386	1,7368
		logístico bicompartimental		0,0776	67,0934	72,7688
	MCMC	exponencial	$N(0; 0, 2^2)$	***	***	***
		logístico		***	***	***
		logístico bicompartimental		0,0442	61,1212	66,7966

2.3.2 Conjunto de dados baseados na produção de gás CO_2 *in vitro*

Nessa seção foram divididas em quatro partes, a primeira apresentando os resultados das análises de *bootstrap*, a segunda apresentando os resultados do MCMC juntamente com a análise frequentista, a terceira comparando as matrizes de correlações dos parâmetros ajustados pelo *bootstrap* e MCMC e finalmente, a quarta parte apresentando a previsão da produção de gases CO_2 por meio de séries temporais.

Resultados *bootstrap*

As estimativas dos parâmetros para as funções não-lineares exponencial, logístico e logístico bicompartimental com 2.000 reamostragem, Tabela 2.7, sobreporam os intervalos de credibilidades apresentados pelas estimativas MCMC (Tabela 2.10) apresentadas na segunda parte dessa seção. O teste de normalidade de Kolmogorov-Smirnov mostra normalidade para distribuição dos parâmetros sobre a forma analítica.

Ajustes usando EDO para a equação logístico bicompartimental teve cerca de 50% de não convergência do algoritmo Levenberg-Marquardt (tentativas de ajustes com outros algo-

Tabela 2.7 – Intervalos de confiança para as médias dos parâmetros com 95% de confiança e teste Kolmogorov-Smirnov (KS) para os modelos não-lineares por análise *bootstrap* com 2.000 reamostragens - produção de gases CO_2 com fermentação *in vitro*

			Bootstrap				
Forma	Modelo	Parâmetros	Média	Intervalo de confiança	Desvio padrão	Teste de KS	Sobreposição Bootstrap e MCMC
			Bootstrap	(inferior; superior)			
Analítica	exponencial	θ_1	1,1096	(1,1084; 1,1108)	0,0006	0,9541	62,28%
		θ_2	0,0955	(0,0946; 0,0965)	0,0004	0,9986	52,63%
		θ_3	-3,9482	(-4,0278; -3,8699)	0,0408	0,5636	56,74%
	logístico	θ_1	1,0879	(1,0852; 1,0904)	0,0013	0,5151	67,92%
		θ_2	0,0384	(0,0373; 0,0394)	0,0005	0,5456	33,33%
		θ_3	-9,8403	(-10,3212; -9,4066)	0,2446	0,6387	38,53%
	bicompartimental	θ_1	0,7828	(0,7742; 0,7915)	0,0042	0,5983	46,82%
		θ_2	0,0646	(0,0631; 0,0660)	0,0008	0,2117	36,55%
		θ_3	-5,8035	(-5,9780; -5,600)	0,0952	0,4131	33,73%
		θ_4	0,3395	(0,3322; 0,3471)	0,0037	0,6034	49,66%
		θ_5	0,0246	(0,0241; 0,0250)	0,0002	0,2447	55,55%
EDO	bicompartimental	θ_1	0,9455	(0,6426; 1,3325)	0,1832	0,3214	1,31%
		θ_2	0,0627	(0,0513; 0,0760)	0,0069	0,0001	2,42%
		θ_3	-7,4826	(-11,7017; -3,5336)	2,1695	0,0693	2,10%
		θ_4	0,3626	(0,3303; 0,3950)	0,0158	0,0450	9,89%
		θ_5	0,0241	(0,0206; 0,0272)	0,0017	0,1032	5,00%

ritmos de otimização, como o Nelder-Mead foram testados e tiveram resultados aproximados). Quando se aplica EDO, em muitos problemas físicos, o interesse é encontrar uma solução particular que satisfaça uma condição inicial do tipo $y(t_0) = y_0$. Esta é a chamada condição inicial e o problema de encontrar uma solução da equação que satisfaça uma condição inicial é denominado de “problema de valor inicial” (Stewart, 2011). Para cada reamostragem de *bootstrap*, o critério adotado para solucionar o problema de valor inicial foi usar o valor mínimo da produção de gases para cada reamostra retirada da amostra original. Ajustes usando a metodologia frequentista e *bootstrap* com o método de Levenberg-Marquardt, para os modelos exponencial e logístico na forma de EDO não foram possíveis de serem ajustados devido a matriz de covariância ser singular.

A média, o desvio padrão (SD), o coeficiente de determinação da regressão linear (R^2), o erro quadrado médio de predição (EQMP), o viés em relação a média (MB), o fator de eficiência da modelagem (MEF), o coeficiente de determinação do modelo (CD), o coeficiente de precisão do modelo (C_b), o valor p para as respectivas hipóteses $H_0: \beta_0 = 0$, $H_0: \beta_1 = 1$ e $H_0: \beta_0 = 0$ e $\beta_1 = 1$, CCC, AIC, AIC_c e o BIC foram usados para comparar os ajustes de compostos de feno de alfafa, Tabela 2.8. A avaliação dos modelos, sobre a metodologia *bootstrap* mostrou que a equação logística bicompartimental na forma analítica foi a que melhor se ajustou ao conjunto das mensurações de gases CO_2 *in vitro*.

Tabela 2.8 – Avaliação dos modelos não-lineares para estimativas de emissões de gases CO_2 no período de 48 horas para feno de alfafa - análise das estimativas de *bootstrap* com 2.000 reamostragens

Estatísticas	Obs.	<i>Bootstrap</i>			
		Analítica			EDO
		exponencial	logístico	logístico bicompartimental	logístico bicompartimental
Média	0,9451	0,9451	0,9455	0,9451	0,9409
SD	0,1912	0,1910	0,1886	0,1912	0,1919
Mediana	1,0210	1,0329	1,0455	1,0240	1,0199
r^2	-	0,9980	0,9910	1,0000	0,9990
EQMP	-	0,0001	0,0003	0,0000	0,0001
MB	-	0,0000	-0,0004	0,0000	0,0042
MEF	-	0,9980	0,9900	1,0000	0,9980
CD	-	1,0020	1,0270	1,0000	0,9920
C_b	-	1,0000	1,0000	1,0000	1,0000
CCC	-	0,9990	0,9950	1,0000	0,9990
valor p^1	-	0,9814	0,0308	0,6402	0,0001
valor p^2	-	0,9870	0,0346	0,6356	0,0024
valor p^3	-	0,9991	0,0977	0,8939	0,0001
AIC_c	-	-5489,5440	-4594,3110	-6297,9380	-5857,7770
AIC	-	-3850,089	-2952,894	-4658,998	-4218,3430
BIC	-	-3837,0370	-2941,8040	-4645,4300	-4205,2690

valor p^1 : (H_0 : $\beta_0 = 0$); valor p^2 : (H_0 : $\beta_1 = 1$) e valor p^3 : (H_0 : $\beta_0 = 0$ e $\beta_1 = 1$).

Resultados das análises MCMC e frequentista

A análise da produção de CO_2 foi ajustada com base nas estimativas do algoritmo de otimização Levenberg-Marquard usando a metodologia frequentista e nas estimativas médias dos parâmetros baseadas nas cadeias MCMC usando a metodologia bayesiana, considerando as três equações não-lineares: exponencial, logístico e logístico bicompartimental.

Verificou-se que o critério de Geweke não apresentou evidência contra a convergência dos parâmetros dos modelos (com nível de 5% de confiança, Tabela 2.9). Com relação ao critério de Raftery e Lewis observou-se que os fatores de dependência para cada parâmetro dos modelos são inferiores a “cinco”.

Os ajustes pelos métodos *bootstrap* (apresentados anteriormente), frequentista e MCMC foram muito próximos para os modelos exponencial e logístico na forma analítica. Os intervalos de confiança e de credibilidade, juntamente com os melhores ajustes das respectivas metodologias frequentista e MCMC foram apresentados na Tabela 2.10, para ambas as formas: analítica e EDO. As distribuições *a posteriori* dos parâmetros foram praticamente coincidentes com as das distribuições *a priori* (Tabela 2.2) apresentada na seção 2.2.6 da metodologia.

Avaliações dos modelos para comparar os ajustes de compostos de feno de alfafa para os ajustes na forma analítica e EDO foram apresentadas nas Tabelas 2.11 e 2.12, respectivamente. A metodologia MCMC foi mais apropriada ao descrever a produção de gás CO_2 dos

Tabela 2.9 – Fator de dependência do critério de Raftery e Lewis (FD) e valor p do critério de Geweke dos modelos não-lineares - produção de gases CO_2 *in vitro*

Modelo	Parâmetro	Forma			
		Analítica		EDO	
		FD	Geweke (valor p)	FD	Geweke (valor p)
exponencial	θ_1	1,0000	0,5238	1,0789	0,6624
	θ_2	1,0526	0,6920	1,1842	0,3197
	θ_3	1,0526	0,3030	1,0526	0,6057
	σ^2	1,0000	0,1916	1,0526	0,7487
logístico	θ_1	1,0789	0,6624	0,9916	0,1202
	θ_2	1,1842	0,3197	0,9518	0,7008
	θ_3	1,0526	0,6057	0,9518	0,3301
	σ^2	1,0526	0,7487	1,0559	0,6465
logístico bicompartimental	θ_1	0,9920	0,6985	1,6400	0,2006
	θ_2	0,9920	0,8661	1,2200	0,5795
	θ_3	0,9920	0,9664	1,0500	0,1943
	θ_4	1,0300	0,6272	1,3200	0,3730
	θ_5	0,9920	0,3778	1,5000	0,9818
	σ^2	0,9920	0,7113	1,0300	0,1989

Tabela 2.10 – Resultados da comparação dos modelos não-lineares por análise frequentista e MCMC na forma analítica e EDO - *prioris* normais - produção de gases CO_2 com fermentação *in vitro*

Modelo			Frequentista		MCMC				
			Média Freq.	Intervalo de confiança (inferior; superior)	Média MCMC	Intervalo de credibilidade (inferior; superior)	EPMC	% aceitação	Sobreposição
Forma Analítica	exponencial	θ_1	1,1096	(1,1088; 1,1104)	1,1096	(1,1088; 1,1106)	0,0000	84,56%	88,88%
		θ_2	0,0955	(0,0951; 0,0960)	0,0955	(0,0950; 0,0960)	0,0000		99,99%
		θ_3	-3,9481	(-3,9843; -3,9119)	-3,9495	(-3,9937; -3,9039)	0,0006		98,28%
	logístico	θ_1	1,0878	(1,0865; 1,0892)	1,0878	(1,0860; 1,0896)	0,0000	84,87%	99,99%
		θ_2	0,0384	(0,0381; 0,0388)	0,0384	(0,0381; 0,0389)	0,0000		99,99%
		θ_3	-9,8277	(-9,8277; -9,6797)	-9,8233	(-10,0000; -9,6476)	0,0018		43,67%
	logístico bicompartimental	θ_1	0,3874	(0,7761; 0,7906)	0,7830	(0,7789; 0,7870)	0,0000	81,80%	55,86%
		θ_2	0,0653	(0,0637; 0,0652)	0,0645	(0,06404; 0,0651)	0,0000		06,75%
		θ_3	-5,8128	(-5,8908; -5,7347)	-5,8084	(-5,8715; -5,7440)	0,0000		86,85%
		θ_4	0,3390	(0,3325; 0,3454)	0,3393	(0,3357; 0,3431)	0,0000		57,36%
		θ_5	0,0245	(0,0205; 0,0284)	0,0245	(0,0243; 0,0248)	0,0000		06,32%
Forma EDO	exponencial	θ_1	*** ¹	*** ¹	1,1090	(1,1080; 1,1100)	0,0000	80,12%	-
		θ_2	*** ¹	*** ¹	0,0975	(0,0970; 0,0980)	0,0000		-
		θ_3	*** ¹	*** ¹	-3,9495	(-4,0272; -3,8679)	0,0009		-
	logístico	θ_1	*** ¹	*** ¹	1,0940	(1,0914; 1,0964)	0,0000	74,06%	-
		θ_2	*** ¹	*** ¹	0,0455	(0,0450; 0,0460)	0,0000		-
		θ_3	*** ¹	*** ¹	-6,8552	(-6,9488; -6,7584)	0,0011		-
	logístico bicompartimental	θ_1	0,8077	(0,7697; 0,8457)	0,8078	(0,8033; 0,8124)	0,0000	81,62%	11,97%
		θ_2	0,0657	(0,0630; 0,0683)	0,0657	(0,0654; 0,0660)	0,0000		11,32%
		θ_3	-5,9927	(-6,5372; -5,4468)	-5,9937	(-6,0815; -5,9071)	0,0011		15,99%
		θ_4	0,3529	(0,3441; 0,3616)	0,3528	(0,3496; 0,3560)	0,0000		36,57%
		θ_5	0,0249	(0,0241; 0,02577)	0,0249	(0,0248; 0,0251)	0,0000		18,75%

¹ Não foi possível estimar covariância, o sistema é singular.

ruminantes *in vitro* em fermentações anaeróbicas tanto para situações na forma analítica como na forma EDO. O melhor ajuste foi obtido com a equação não-linear logística bicompartimental e a solução na forma analítica produziu melhores resultados do que a forma EDO (menores AIC_c , menores AIC e menores BIC), Tabela 2.12. O ajuste do valor inicial, y_0 , na EDO é tomado como fixo e no modelo analítico, o valor inicial é aleatório, o que pode ocasionar uma menor flexibilidade no ajuste da curva, além do algoritmo numérico, *lsoda*, usado para aproximar a solução da equação.

Mello et al. (2008) ajustaram diferentes equações não-lineares na produção acumulativa de gases em silagens de girassol e milho usando a técnica *in vitro*, em que as leituras

Tabela 2.11 – Avaliação dos modelos não-lineares para estimativas de emissões de gases CO_2 no período de 48 horas para feno de alfafa - forma analítica

Estatísticas	Obs.	Frequentista			MCMC		
		exponencial	logístico	bicompartmental	exponencial	logístico	bicompartmental
Média	0,9451	0,9449	0,9454	0,9451	0,9451	0,9454	0,9451
SD	0,1912	0,1910	0,1887	0,1912	0,1910	0,1887	0,1912
Mediana	1,0210	1,0327	1,0456	1,0240	1,0329	1,0456	1,0240
r^2	-	0,9980	0,9910	1,0000	0,9980	0,9910	1,0000
EQMP	-	0,0001	0,0003	0,0000	0,0001	0,0003	0,0000
MB	-	0,0002	-0,0003	0,0000	0,0000	-0,0003	0,0000
MEF	-	0,9980	0,9900	1,0000	0,9980	0,9900	1,0000
CD	-	1,0010	1,0260	1,0000	1,0020	1,0260	1,0000
C_b	-	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
CCC	-	0,9990	0,9950	1,0000	0,9990	0,9950	1,0000
valor p^1	-	0,0000	0,0000	0,60951	0,9818	0,0395	0,6056
valor p^2	-	0,8540	0,0451	0,61549	0,9795	0,0439	0,6143
valor p^3	-	0,8077	0,1242	0,8775	0,9996	0,1208	0,8754
AIC_c	-	-5489,5100	-4593,8300	-6298,0290	-5489,5440	-4593,8980	-6298,0350
AIC	-	-3850,1000	-2954,4000	-4658,570	-3850,1100	-2954,4640	-4658,6000
BIC	-	-3837,0300	-2941,3200	-4645,496	-3837,0370	-2941,3900	-4645,5180

valor p^1 : ($H_0: \beta_0 = 0$); valor p^2 : ($H_0: \beta_1 = 1$) e valor p^3 : ($H_0: \beta_0 = 0$ e $\beta_1 = 1$).

Tabela 2.12 – Avaliação dos modelos não-lineares para estimativas de emissões de gases CO_2 no período de 48 horas para feno de alfafa - EDO

Estatísticas	Obs.	Frequentista			MCMC		
		Exponencial	logístico	Logístico bicompartmental	exponencial	logístico	logístico bicompartmental
Média	0,9451	-	-	0,9447	0,9415	0,9326	0,9447
SD	0,1912	-	-	0,1923	0,1952	0,2070	0,1923
Mediana	1,021	-	-	1,0240	1,0312	1,0468	1,0240
r^2	-	-	-	0,9990	0,9980	0,9860	0,9990
EQMP	-	-	-	0,0000	0,0001	0,0009	0,0000
MB	-	-	-	0,0004	0,0036	0,0125	0,0004
MEF	-	-	-	0,9990	0,9970	0,9740	0,9990
CD	-	-	-	0,9890	0,9590	0,8500	0,9880
C_b	-	-	-	1,0000	1,0000	0,9950	1,0000
CCC	-	-	-	1,0000	0,9990	0,9880	1,0000
valor p^1	-	-	-	0,0000	0,0000	0,0000	0,0000
valor p^2	-	-	-	0,0000	0,0000	0,0000	0,0000
valor p^3	-	-	-	0,0000	0,0000	0,0000	0,0000
AIC_c	-	-	-	-6194,377	-5489,5440	-4385,0220	-6194,737
AIC	-	-	-	-4554,943	-3850,1100	-2745,5880	-4555,3030
BIC	-	-	-	-4541,87	-3837,0370	-2732,5140	-4542,230

valor p^1 : ($H_0: \beta_0 = 0$); valor p^2 : ($H_0: \beta_1 = 1$) e valor p^3 : ($H_0: \beta_0 = 0$ e $\beta_1 = 1$)

de pressões (em volts) foram medidas no período de tempo entre 1 até 192 horas após o período de incubação. Seus resultados mostraram que usando os ajustes na forma analítica de equações não-lineares Brody e France subestimam a produção de gases como valores negativos, nos tempos iniciais de incubação (0 a 3 horas) o que é biologicamente incoerente. Pell et al. (2000) citado por Mello et al. (2008) mencionaram que leituras mais frequentes na fase inicial do modelo ($n \geq 16$) são suficientes para proporcionar um ajuste mais adequado.

De acordo com Shofield et al. (1994) e Groot et al. (1996), os modelos multicompartimentais apresentam maior qualidade de ajuste que os modelos baseados em cinética de primeira ordem. Segundo Fondevila e Barrios (2001) os modelos multicompartimentais são mais adequados quando se utilizam um número muito grande de leituras, como o caso desse presente estudo com 577 observações. Os traços das cadeias e o melhor ajuste foi representado pela equação logística bicompartmental, respectivamente, nas Figuras 2.2 e 2.3. As estimativas quanto a

produção acumulativa de gases (CO_2), atingiu o valor máximo de 1,1099 (mL) no tempo de 48 h .

A precisão das equações depende em grande medida dos parâmetros que descrevem as características relevantes da produção de CO_2 que representa o sistema digestivo dos ruminantes. Alguns modelos de nutrição podem representar um grande número de parâmetros de entrada e portanto, não é possível estudar o ajuste simultaneamente de todos eles na função. É importante identificar os parâmetros que mais influenciam o modelo e sua importância nos resultados gerados, ou seja, a intensidade que faz com que cada mudança no modelo, de modo que podemos dar mais ênfase à aquisição e refinamento destes parâmetros e assim, reduzir a incerteza e aumentar a fiabilidade dos resultados obtidos (JHA, 2009). A importância dos parâmetros é fornecida pelo valor absoluto da função sensibilidade. O modelo logístico bicompartimental teve maior influência dos parâmetros θ_1 , θ_2 e θ_3 , enquanto que θ_4 e θ_5 são menos influentes (Figura 2.4).

Figura 2.2 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 , θ_3 , θ_4 e θ_5) na forma EDO Logístico bicompartimental - produção de gases CO_2 fermentação de feno de alfafa experimento *in vitro*

Mello et al. (2008) mostraram que o modelo logístico bicompartimental apresentou maior qualidade de ajuste na produção acumulativa de gases em silagens de girassol e milho usando a técnica *in vitro*. Nossos resultados apresentaram valores menores de EQMP comparados com as curvas ajustadas para os compostos de silagens de girassol e milho apresentados por esses autores. Os valores de r^2 apresentados por esses autores em relação a equação logística bicompartimental foram 0,9951 e 0,9954, respectivamente para os compostos de girassol e milho. Resultados semelhantes foram encontrados por Schofield et al. (1994), $r^2 > 0,999$, na produção *in vitro* de substratos com celulosas. Apesar disso, as diferenças entre os modelos e substratos foram irrisórias e não tem significado prático, pois o r^2 não é um critério adequado para verificar a qualidade do ajustamento. Mello et al. (2008) afirmam que o r^2 fornece apenas

Figura 2.3 – Ajuste MCMC da equação não-linear Logística bicompartimental - produção de gases (CO_2) fermentação de feno de alfafa experimento *in vitro*

Figura 2.4 – Sensibilidades dos parâmetros do Logístico bicompartimental - produção de gases (CO_2) fermentação de feno de alfafa experimento *in vitro*

uma idéia da associação de $\hat{y}$ em relação a y , mas pouco alega sobre a aproximação biológica da curva, principalmente em relação a sua forma.

O método RAMA no pacote FME para geração de realizações de variáveis aleatórias com a estrutura de uma matriz de covariância multivariada dos parâmetros mostrou resultados promissores para solução de EDO e equações não-lineares. Estudos com simulações comparando convergências com o uso do método MCMC simples e a implementação do RAMA, mostram que esse último método fornece um diagnóstico sistemático para remediar o processo de adaptação da cadeia, Haario et al. (2006).

Como a implementação desses métodos é recente no *software* livre R, futuras contribuições ainda serão acrescentadas, como descritos pelos autores do pacote, a ampliação da função para trabalhar com mais de uma variável dependente (ex.: tempo e espaço). A aplicação de outros modelos não-lineares na área médica (modelo que simula a dinâmica do vírus HIV) e na hidrologia (modelo para simular a água e ciclagem de nutrientes em bacias hidrográficas) tem sido calibrado com esta ferramenta, com uso de *prioris* não informativas (Soetaert e Petzoldt, 2010; Wu e Liu, 2012). Os autores mencionam que a ferramenta apresenta um bom desempenho na procura de conjuntos ótimos de parâmetros.

Wu et al. (2015) fizeram um estudo sobre a relação entre o consumo de energia, a população urbana, o crescimento econômico e as emissões de CO_2 nos países: Brasil, Rússia, Índia, China e África do Sul. Seus resultados indicam que o crescimento econômico tem um efeito de diminuição nas emissões de CO_2 no Brasil e na Rússia e tem um efeito crescente na Índia, China e África do Sul. As emissões de CO_2 aumentarão com o aumento da população e o consumo de energia urbana no Brasil, Rússia, Índia, China e África do Sul.

As emissões de gases de efeito estufa e o aquecimento global tornaram-se as questões de investigação importantes nos campos da ciência e da política. Os estudos existentes indicam que mais de dois terços dos gases de efeito estufa provenientes de emissões de CO_2 estão relacionados com a energia fóssil (Wu et al., 2015). Assim, a previsão das emissões de CO_2 e análise de seus fatores de influência são fundamentais para a adaptação das políticas e diminuição das consequências da mudança climática. A previsão de emissões de CO_2 tem sido amplamente discutida em diversos setores.

Os principais limites para a produção de gado nos países em desenvolvimento é a quantidade escassa e qualidade do fornecimento de alimentação durante todo o ano. Estes países enfrentam grave escassez de alimentos para animais do tipo convencional. Alimentos como grãos são necessários quase exclusivamente para consumo humano. A população mundial está a crescer a uma taxa muito elevada e a maior parte do crescimento da população está ocorrendo em países em desenvolvimento, como no Brasil (Makkar, 2002).

Com o aumento da demanda por produtos animais, resultado do rápido crescimento das economias mundiais e diminuição das áreas de terra, esperanças futuras de alimentar as milhares de pessoas e salvaguardar a sua segurança alimentar vai depender da melhor utilização dos recursos alimentares não convencionais, que não pode ser utilizado como alimento para seres humanos. Além disso, uma grande área de terra no mundo está degradada, infértil ou são reservas (margens de rios) e a quantidade de terras improdutivas está aumentando a cada ano (Makkar, 2002).

As regiões tropicais caracterizam-se pelo grande número de espécies forrageiras com potencial para serem usadas na alimentação de ruminantes e na determinação do seu valor nutritivo. Uma revisão de literatura levantada com 24 artigos relacionados com produção de gases no Brasil, com mais variados tipos de compostos de dietas, com diferentes localizações, altitudes, pressão atmosférica, tempo de coleta e inóculo ruminal (publicados a partir do ano 2000 em periódicos nacionais e internacionais), mostram que os ajustes da produção de gases em sua maioria são ajustados com o modelo logístico bicompartimental ($\simeq 41\%$) e o modelo France ($\simeq 29\%$), Tabela 2.13. Assim, o uso de metodologias via reamostragem por *bootstrap* não-paramétrico e de métodos bayesianos implementados no pacote FME podem ser uma ferramenta útil para obter uma melhora nas estimativas e estudo de incerteza dos parâmetros para os modelos não-lineares, tanto na forma analítica como sistema e equações diferenciais e balancear melhores dietas tentando minimizar fermentações inúteis que produzam GEE.

Modesto et al. (2004), ao avaliarem silagens do terço superior da rama da mandioca, obtiveram, estimativas para os parâmetros do modelo logístico bicompartimental θ_2 e θ_5 , 0,123 e 0,025 h^{-1} , respectivamente. Valores similares de θ_5 (0,0245 h^{-1}) foi obtido neste estudo para o composto de feno de alfafa, porém inferior para a θ_2 (0,0645 h^{-1}). Conclusão semelhante pode ser observada para silagem de parte área de mandioca sem acréscimo de raízes, para θ_2 e θ_5 , 0,084 e 0,084 h^{-1} , respectivamente, em Silva et al. (2012) em relação aos resultados de Modesto et al. (2004). Todos esses estudos (Modesto et al. (2004), Silva et al (2012) e o do apresentado

nesse trabalho), apresentaram valores aquém dos relatados por Campos et al. (2000) para o parâmetro θ_5 em silagem de milho e feno de alfafa, respectivamente, 0,042 e 0,037 h^{-1} . A menor taxa de digestão da fração fibrosa da silagem da parte aérea de mandioca e de feno de alfafa (desse presente estudo) é provavelmente devido à elevada concentração de lignina em sua FDN, o que limita o acesso microbiano, retardando a degradação dos compostos da parede celular.

Tomich et al. (2003) ressaltaram a importância dos parâmetros cinéticos como critério de seleção ou avaliação da qualidade das forragens, não devendo, contudo ser levados em consideração individualmente, pois maiores valores para θ_1 e θ_4 (valor de potencial máximo de produção de gases) indicam maior fermentabilidade do material, sem considerar a limitação do tempo de permanência do alimento no rúmen, enquanto os altos valores de θ_2 e θ_5 (valor da taxa de produção de gases) implicam menor tempo de degradação para as frações potencialmente fermentescíveis (digestíveis), mas desconsideram a extensão desta fração no alimento. As forragens mais digestíveis, seriam aquelas que apresentam altos valores de potencial máximo e taxa de produção de gases, pois alcançariam potenciais máximos de fermentação em menor tempo.

Segudo Silva et al. (2012), a cinética de digestão e de passagem do alimento pode prover dados sobre a eficiência de utilização dos alimentos, fornecendo parâmetros importantes para avaliação e seleção dos alimentos. Maiores taxas de degradação poderão resultar em aumento da velocidade de fluxo da digestão, elevando a ingestão de matéria seca e a quantidade de substratos disponíveis para a biossíntese microbiana, o que acarretará maior oferta de proteína bruta no intestino delgado e, por conseguinte, melhor desempenho animal.

Tabela 2.13 – Artigos acadêmicos que retratam produção de gases de efeito estufa na agropecuária no Brasil com publicações desde o ano 2000

Referência	Tipo de técnica	Compostos	Modelo Estatístico	Período da coleta (horas) após incubação
Cabral et al. (2000)	<i>in vitro</i>	capim-elefante; capim tifton 85; fenos de alfafa e de capim-coastcross; SM ¹³ e SS ¹⁴ ; e a cana-de-açúcar	logístico unicompartimental	1/2, 1, 2, 3, 4, 5, 6, 9, 12, 18, 24, 30, 36, 48, 60 e 72
Campos et al. (2000)	<i>in vitro</i> e <i>in situ</i>	SM ¹³ com alto/baixo teor de matéria seca com/sem inoculante	logístico bicompartimental	0 até 48
Gonçalves et al. (2001)	<i>in vitro</i> e <i>in situ</i>	feno de alfafa; capim-elefante; feno de Tifton 85 e SM ¹³	logístico unicompartimental	1/2, 1, 2, 3, 4, 5, 6, 9, 12, 18, 24, 30, 36, 48, 60 e 72
Fernandes et al. (2003)	<i>in vitro</i>	cana-de-açúcar	logístico bicompartimental	0, 12, 24, 36, 48, 60, 72, 84 e 96
Maurício et al. (2003)	<i>in vitro</i> semiautomática	tifton 85; braquiário; SS ¹⁴ ; BR700, BR701, BR601 e AG2002	Equação quadrática relacionando volume e pressão	2, 4, 6, 8, 10, 12, 15, 19, 24, 30, 36, 48, 72 e 96
Tomich et al. (2003)	<i>in vitro</i>	híbridos de sorgo com capim-sudão	France	2, 4, 6, 8, 10, 12, 15, 19, 24, 30, 36, 48, 72 e 96
Modesto et al. (2004)	<i>in vitro</i>	silagem da rama de mandioca	logístico bicompartimental	1, 2, 3, 4, 5, 6, 8, 10, 12, 15, 24, 27, 30, 33, 36, 48, 56, 60, 72, 96 e 120
Nogueira et al. (2006)	<i>in vitro</i> semiautomática	cana-de-açúcar, SM ¹³ , e capim-colômbio, milho em grão ração comercial para vacas em lactação	France	2, 4, 6, 8, 10, 12, 15, 19, 24, 30, 36, 48, 72 e 96
Pereira et al. (2006)	<i>in vitro</i>	45% de SM ¹³ e 55% de concentrado composto por milho; farelo de girassol; casca de soja; farelo de soja	delineamento em blocos (períodos) em esquema fatorial (3 × 5)	quatro períodos diferentes não especificados
Castro et al. (2007)	<i>in vitro</i> semiautomática	capim-braquiário cortado aos 28, 56, 84 e 112 dias	France	2, 4, 6, 8, 10, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72 e 96
Guimarães Jr. et al. (2008)	<i>in vitro</i> semiautomática	genótipos de milho: BRS-1501, NPM-1 e CMS-3	France	2, 4, 6, 8, 10, 12, 15, 19, 24, 30, 36, 48, 72 e 96
Mello et al. (2008)	<i>in vitro</i>	silagem de girassol e SM ¹³	Brody, Von Bertalanffy, Gompertz, France, logístico, logístico modificado e logístico bicompartimental	1, 2, 3, 4, 5, 6, 9, 12, 18, 24, 30, 36, 48, 60, 72, 84, 96, 120, 144, 168 e 192
Detmann et al. (2009)	<i>in vitro</i>	capins setária, hemarthria, angola e acroceres, adubados com 0, 100, 200, 300 ou 400 kg de N/ha e colhidos aos 28, 42, 56 ou 70 dias de idade.	logístico bicompartimental	0, 1, 2, 3, 4, 5, 6, 8, 10, 12, 18, 24, 30, 36, 48, 60, 72, 96, 120 e 144
Moreira et al. (2010)	<i>in vitro</i> semiautomática	CEDE ¹ ; CEDU ² ; CEPC ³ ; CADE ⁴ ; CADU ⁵ ; CAPC ⁶ ; SMDE ⁷ ; SMDU ⁸ ; SMP ⁹ ; SRDE ¹⁰ ; SRDU ¹¹ ; SRPC ¹² .	France	2, 6, 12, 24, 48, e 96
Wascheck et al. (2010)	<i>in vitro</i>	Farelo de arroz parboilizado; Farelo de arroz integral; 50% Fubá de milho e 50% Farelo de arroz integral; 20% Fubá de milho e 80% Farelo de arroz integral; 50% Fubá de milho e 50% Farelo de arroz parboilizado; 20% Fubá de milho e 80% Farelo de arroz parboilizado Fubá de milho;	France	3, 6, 8, 10, 12, 15, 18, 21, 24, 27, 30, 36, 48, 72 e 96
Farias et al. (2011)	<i>in vitro</i> semiautomática	farelo e torta de babaçu	logístico bicompartimental e Gompertz	3, 6, 9, 12, 15, 21, 27, 33, 39, 48, 60, 72 e 96
Fortaleza et al. (2011)	<i>in vitro</i> semiautomática	rações contendo 0, 2, 4 e 8 % de inclusão de torta de nabo forrageiro	Gompertz	1/2, 1, 2, 4, 6, 8, 10, 12, 15, 18, 21, 24, 30, 36 e 48
Mizubuti et al. (2011)	<i>in vitro</i>	farelo de crumbe; farelo de algodão; torta de crumbe; torta de soja e torta de girassol.	logístico bicompartimental	1, 2, 3, 4, 5, 6, 9, 12, 18, 24, 30, 36, 48, 60, 72, 84, 96 e 144
Rodrigues et al. (2011)	<i>in vitro</i> semiautomática	silagens de maniçoba com diferentes das inclusões do coproduto de vitivinícolas desidratado (CVD), distribuídos em quatro tratamentos (0, 8, 16 e 24% de inclusão do CVD)	logístico bicompartimental	2, 5, 6, 8, 9, 11, 12, 14, 17, 20, 24, 28, 34, 48, 72 e 96
Sá et al. (2011)	<i>in vitro</i>	Brachiaria brizantha cv. Marandu	logístico bicompartimental	2, 4, 6, 8, 10, 12, 14, 17, 20, 24, 28, 32, 48, 72 e 96
Araújo et al. (2012)	<i>in vitro</i>	capins buffel; gramão; pangola; brachiaria	regressão linear	6, 24 e 48 horas
Silva et al. (2012)	<i>in vitro</i>	silagens da parte aérea da mandioca inclusão 0, 15, 30 e 45% de raízes da mandioca	logístico bicompartimental	6, 12, 24, 48 e 96
Morgado et al. (2013)	<i>in vitro</i>	casca de soja, farelo de girassol, milho, polpa cítrica e SM ¹³	delineamento inteiramente casualizado em medidas repetidas no tempo	quatro períodos diferentes após 12 horas de fermentação
Morais et al. (2015)	<i>in vitro</i> semiautomática	capim-elefante com uma idade de corte de 50 dias subprodutos da indústria de biodiesel: mamona, algodão, moringa, Pinhão manso	regressão linear	0 até 48

CEDE¹: capim elefante + milho dentado, CEDU²: capim elefante + milho duro, CEPC³: capim elefante + polpa cítrica, CADE⁴: cana-de-açúcar + milho dentado, CADU⁵: cana-de-açúcar + milho duro, CAPC⁶: cana-de-açúcar + polpa cítrica, SMDE⁷: silagem de milho + milho dentado, SMDU⁸: silagem de milho + milho duro, SMP⁹: silagem de milho + polpa cítrica, SRDE¹⁰: silagem de milho + polpa cítrica, SRDU¹¹: silagem de milho + polpa cítrica, SRPC¹²: silagem de milho + polpa cítrica, SM: silagem de milho¹³; SS¹⁴: silagem de sorgo.

Comparação das estruturas de correlações entre as metodologias *bootstrap* e MCMC

Nessa seção, foi apresentado a comparação pelo teste de aleatorização de Mantel das distâncias das estruturas de correlações entre os ajustes realizados por *bootstrap* e MCMC. As Figuras 2.5, 2.6, 2.7 e 2.8 apresentam as matrizes de dispersão conjunta com os histogramas de frequência dos ajustes na forma analítica exponencial, logística, logística bicompartimental e na forma de EDO da logística bicompartimental, respectivamente, para as duas metodologias (*bootstrap* à esquerda e MCMC à direita). A direção das dispersões dos parâmetros mostram um boa correspondência do sentido da correlação de *Pearson* entre os resultados de *bootstrap* e MCMC.

Figura 2.5 – Matriz de dispersão conjunta com os histogramas de frequência dos ajustes *bootstrap* (à esquerda) e ajustes MCMC (à direita) para produção de gases *in vitro* no período de 48 horas, modelo exponencial na forma analítica

Figura 2.6 – Matriz de dispersão conjunta com os histogramas de frequência dos ajustes *bootstrap* (à esquerda) e ajustes MCMC (à direita) para produção de gases *in vitro* no período de 48 horas, equação logística na forma analítica

Figura 2.7 – Matriz de dispersão conjunta com os histogramas de frequência dos ajustes *bootstrap* (à esquerda) e ajustes MCMC (à direita) para produção de gases *in vitro* no período de 48 horas, equação logística na forma analítica

Figura 2.8 – Matriz de dispersão conjunta com os histogramas de frequência dos ajustes *bootstrap* (à esquerda) e ajustes MCMC (à direita) para produção de gases *in vitro* no período de 48 horas, equação logística na forma de EDO

O teste aleatorização de Mantel confirma que as distâncias entre as correlações dos parâmetros das equações estimadas pelas metodologias *bootstrap* e MCMC são iguais ao nível de 1% de confiança, Tabela 2.14.

Apesar das três metodologias apresentarem convergências nos resultados, a análise residual mesmo do melhor ajuste (equação logística bicompartimental) apresentou algumas complicações como dependência dos resíduos (visto que o conjunto de dados são cumulativos ao longo do tempo) e heterogeneidade de variância. Os resíduos apresentaram normalidade e as autocorrelações foram ajustadas aos modelos não-lineares usando a função *glns*, mas não teve melhora significativa na análise residual. A abordagem da produção de gases *in vitro* foi ajustada por meio de séries temporais que contempla situações como correlações entre as observações no

Tabela 2.14 – Teste aleatorização de Mantel para comparar as matrizes de correlações dos parâmetros dos modelos estimados pela metodologia *bootstrap* e MCMC, com 1.000 permutações

Modelo	Estrutura de correlações nas formas	Estatística calculada	valor p
exponencial	analítica	0,9882	0,0009
logístico	analítica	0,9969	0,0009
logístico bicompartimental	analítica	0,9424	0,0010
logístico bicompartimental	EDO	0,7367	0,0080

instante t e um tempo subsequente $t + h$.

Resultados da Série temporal

Com o objetivo de buscar melhores resultados das análises residuais dos modelos não-lineares (exponencial, logístico e logístico bicompartimental) ajustados no banco de dados de gases de produção *in vitro* usando metodologias de *bootstrap* e MCMC, nesta subseção são apresentados os resultados da produção de gases utilizando a metodologia de séries temporais. Na Figura 2.9 é apresentado o gráfico da série original. Visualmente, pode-se instituir a presença da componente tendência e a não ocorrência da componente sazonal. Mas, para a confirmação da presença ou não destas componentes, deve-se aplicar o Teste de Cox-Stuart e o teste de Fisher, respectivamente, para verificar a existência de tendência e sazonalidade, descritas na seção 1.1.3.

Figura 2.9 – Representação gráfica da série original de produção de gases no período de 47,4167 horas, em mL

A análise espectral para série, representada pelo periodograma, Figura 2.10, fornece os possíveis períodos cíclicos e sazonais da série. Para o período sazonal observa-se a não existência de picos, indicando que a série não possui sazonalidade. A ausência da componente de sazonalidade foi confirmada pelo teste de Fisher.

Figura 2.10 – Periodograma da série de produção de gases no período de 47,4167 horas, em mL

A tendência da série é testada por meio do teste dos sinais de Cox-Stuart considerando um nível de significância $\alpha = 0,05$, testando-se as hipóteses H_0 : não existe tendência e H_1 : existe tendência. Como $N = 570$ observações tem-se duas amostras cada uma com 285 observações. O número de sinais positivos, T_2 , foi igual a 285. Assim,

$$c = \frac{570}{2} = 285, \quad n = 285, \quad p = q = \frac{1}{2} \quad \text{e} \quad t \sim N(np, npq) \quad (2.59)$$

Como $T_2 \geq n - t$, ($T_2 \geq 285 - 156,38 = 128,62$), rejeita-se a hipótese de nulidade, confirmando a presença de tendência.

Identificação

A indentificação do modelo que se ajusta aos dados de uma série temporal foi realizada através das FAC e FACP da série estacionária. O teste dos sinais de Cox-Stuart confirmou a presença da componente de tendência na série e portanto, se faz necessária a aplicação da primeira diferença nesta, a fim que se torne estacionária. A Figura 2.11 apresenta a série estacionária.

Figura 2.11 – Representação gráfica da série diferenciada de produção de gases no período de 47,4167 horas, em mL

As funções FAC e FACP da série diferenciada possuem a característica de identificar as ordens do modelo, apresentadas na Figura 2.12. Por meio da FAC e FACP observa-se que a série não apresenta sazonalidade determinística, isto é, ausência de senóides. A ausência da componente de sazonalidade foi confirmada pelo teste de Fisher.

Figura 2.12 – Função de autocorrelação e autocorrelação parcial da série diferenciada de produção de gases no período de 47,4167 horas, em mL

Os modelos propostos foram o modelo ARIMA(p, d, q):

$$\phi(B) \Delta^d Z_t = \theta(B) a_t \quad \text{ou} \quad \phi(B)(1 - B)^d Z_t = \theta(B) a_t \quad (2.60)$$

em que $\phi(B) = 1 - \phi_1(B) - \phi_2(B^2) - \dots - \phi_p(B^p)$ é o operador autoregressivo de ordem p ; Δ^d é o operador diferença simples, d indicando o número de diferenças; $\theta(B) = \theta_1(B) - \theta_2(B^2) - \dots - \theta_q(B^q)$ é o operador de médias móveis de ordem q ; a_t é o resíduo que, eventualmente, pode ser ruído branco.

Os modelos que apresentaram ruído branco foram:

Modelo 01: ARIMA (1, 1, 1)

$$1 - \phi_1 B(1 - B)^1 Z_t = [1 - \theta_1(B)]a_t \quad (2.61)$$

Modelo 02: ARIMA (1, 1, 2)

$$1 - \phi_1 B(1 - B)^1 Z_t = [1 - \theta_1(B) - \theta_2(B^2)]a_t \quad (2.62)$$

Modelo 03: ARIMA (3, 1, 0)

$$1 - \phi_1(B) - \phi_2(B^2) - \phi_3(B^3)(1 - B)^1 Z_t = a_t \quad (2.63)$$

Modelo 04: ARIMA (4, 1, 0)

$$1 - \phi_1(B) - \phi_2(B^2) - \phi_3(B^3)\phi_4(B^4)(1 - B)^1 Z_t = a_t \quad (2.64)$$

As estimativas dos parâmetros para os quatro modelos propostos na fase de identificação foram apresentados na Tabela 2.15. O método utilizado para a obtenção das estimativas foi o de máxima verossimilhança do software STATISTICA[®].

Tabela 2.15 – Estimativas dos parâmetros dos modelos de séries temporais ajustados aos gases CO_2 em mL , para o período de 48 horas

Modelo	Parâmetro	Estimativa	Erro padrão	valor p	LI	LS
ARIMA(1, 1, 1)	ϕ_1	0,9765	0,0103	0,0000	0,9561	0,9969
	θ_1	0,5934	0,0461	0,0000	0,5028	0,6840
ARIMA(1, 1, 2)	ϕ_1	0,9868	0,0078	0,0000	0,9714	1,0021
	θ_1	0,5172	0,0434	0,0000	0,4318	0,6026
	θ_2	0,1806	0,0477	0,0001	0,0869	0,2743
ARIMA(3, 1, 0)	ϕ_1	0,5231	0,0403	0,0000	0,4439	0,6023
	ϕ_2	0,1113	0,0457	0,0152	0,0214	0,2011
	ϕ_3	0,2866	0,0404	0,0000	0,2072	0,3660
	ϕ_1	0,4941	0,0419	0,0000	0,4118	0,5764
ARIMA(4, 1, 0)	ϕ_2	0,0997	0,0457	0,0297	0,0098	0,1897
	ϕ_3	0,2337	0,0458	0,0000	0,1437	0,3237
	ϕ_4	0,1014	0,0420	0,0160	0,0189	0,1840

As Figuras 2.13, 2.14, 2.15 e 2.16 são as funções de autocorrelação residual e as funções de autocorrelação parcial residual dos modelos propostos. Observando estas funções por meio de uma inspeção visual, pode se concluir que os modelos são adequados, visto que menos de 5% dos “lags” são significativos. Na Tabela 2.16, é mostrado formalmente por meio do Teste de Box-Pierce que todos os resíduos dos modelos propostos constituem ruído branco, portanto,

os modelos se ajustaram bem a produção de gases acumulativa de CO_2 .

Figura 2.13 – Função de autocorrelação residual e autocorrelação parcial residual da série temporal, modelo ARIMA (1, 1, 1)

Figura 2.14 – Função de autocorrelação residual e autocorrelação parcial residual da série temporal, modelo ARIMA (1, 1, 2)

Figura 2.15 – Função de autocorrelação residual e autocorrelação parcial residual da série temporal, modelo ARIMA (3, 1, 0)

Figura 2.16 – Função de autocorrelação residual e autocorrelação parcial residual da série temporal, modelo ARIMA (4, 1, 0)

Tabela 2.16 – Estatísticas para o teste de Box-Pierce de modelos ARIMA

Modelo	GL	Q_{40}	χ^2
ARIMA (1, 1, 1)	38	34,91	53,38
ARIMA (1, 1, 2)	37	27,22	52,19
ARIMA (3, 1, 0)	37	33,63	52,99
ARIMA (4, 1, 0)	36	30,02	53,38

A escolha do melhor modelo, baseada nas estatísticas do AIC, BIC e do EQMP foram apresentadas na Tabela 2.17. Assim, entre os modelos propostos, o modelo ARIMA (1, 1, 2) apresentou o melhor ajuste (menores valores de AIC, BIC e EQME) para realizar as previsões de gases CO_2 .

Tabela 2.17 – Estimativas dos critérios de AIC, BIC e EQMP de modelos ARIMA

Modelo	AIC	BIC	EQMP
ARIMA (1, 1, 1)	-7704,9140	-13,1346	0,0000131
ARIMA (1, 1, 2)	-7715,9773	-13,5665	0,0000085
ARIMA (3, 1, 0)	-7703,9135	-13,1378	0,0000166
ARIMA (4, 1, 0)	-7706,9246	-13,2005	0,0000130

Previsão

Para os quatro modelos que são ruído branco, foram apresentadas as previsões (estimativas para os valores preditos de gases CO_2 mL), os valores reais e o erro de previsão para o período $t = 47,5$ à $t = 48$ horas, Tabela 2.18. Todos os modelos apresentados subestimaram os valores observados de gases CO_2 . O modelo ARIMA (1, 1, 2) foi o que apresentou uma aproximação melhor, conforme já discutido os resultados da Tabela 2.17, quanto o melhor ajuste.

Sua representação algébrica está apresentada na Eq. (2.65):

$$1 - 0,9868B(1 - B)^1 Z_t = [1 - 0,5172(B) - 0,1806(B^2)]a_t \quad (2.65)$$

em que a produção acumulativa de medidas de gases CO_2 , atinge o valor de 1,1054 mL no tempo de 48 h.

Tabela 2.18 – Valores observados de gases CO_2 , para o período de $t = 47,5$ à $t = 48$ horas e valores preditos para o mesmo período

Modelo	tempo	Z_{t+h}	$\hat{Z}_t(h)$	$e_t(h)$
ARIMA (1, 1, 1)	47,5000	1,1080	1,1065	0,0014
	47,5833	1,1080	1,1062	0,0018
	47,6667	1,1080	1,1058	0,0022
	47,7500	1,1090	1,1054	0,0036
	47,8333	1,1090	1,1050	0,0040
	47,9167	1,1090	1,1046	0,0043
	48,0000	1,1100	1,1043	0,0057
ARIMA (1, 1, 2)	47,5000	1,1080	1,1066	0,0013
	47,5833	1,1080	1,1064	0,0015
	47,6667	1,1080	1,1062	0,0017
	47,7500	1,1090	1,1060	0,0029
	47,8333	1,1090	1,1060	0,0032
	47,9167	1,1090	1,1056	0,0034
	48,0000	1,1100	1,1054	0,0046
ARIMA (3, 1, 0)	47,5000	1,1080	1,1066	0,0013
	47,5833	1,1080	1,1061	0,0019
	47,6667	1,1080	1,1054	0,0025
	47,7500	1,1090	1,1050	0,0040
	47,8333	1,1090	1,1044	0,0045
	47,9167	1,1090	1,1040	0,0050
	48,0000	1,1100	1,1035	0,0064
ARIMA (4, 1, 0)	47,5000	1,1080	1,1067	0,0012
	47,5833	1,1080	1,1063	0,0016
	47,6667	1,1080	1,1058	0,0021
	47,7500	1,1090	1,1054	0,0036
	47,8333	1,1090	1,1050	0,0040
	47,9167	1,1090	1,1045	0,0044
	48,0000	1,1100	1,1042	0,0058

A representação gráfica do modelo ARIMA (1, 1, 2) da produção de gás CO_2 para previsão de até 6 horas após o período $t = 47,5$ horas é apresentado na Figura 2.17. Por meio das previsões é possível observar que o maior pico da produção de gases foi no tempo de 47,5 horas, em que atingiu uma produção máxima total de 1,1066 mL. Após esse período houve uma queda gradativa na produção de CO_2 .

Figura 2.17 – Previsão da produção de gases CO_2 *in vitro* após o período de 47,5 horas

2.4 Conclusões

No conjunto de dados sintéticos, simulada por meio de parâmetros fixos e acréscimo de ruídos representando o erro aleatório (desvios padrão iguais a 0,05 e 0,2), as metodologias frequentista e MCMC tiveram comportamento semelhante no ajuste do modelo exponencial para todos os cenários. O método frequentista foi melhor para ajustar o modelo logístico e a metodologia MCMC foi melhor para ajustar o modelo logístico bicompartimental, em todos os cenários estudados (sobre os critérios EQMP, AIC e BIC). No conjunto de dados reais de produção de CO_2 obtidos por meio de fermentação *in vitro* de feno de alfafa, o modelo não-linear logístico bicompartimental na forma analítica apresentou melhor ajuste pela metodologia MCMC em relação a todos os demais modelos (sobre os critérios EQMP, AIC e BIC), mas apresentou problemas de dependência residual. Os modelos de séries temporais estimaram melhor os valores de produção de gases CO_2 . O melhor modelo foi determinado pelo ARIMA (1, 1, 2) em que a produção acumulativa de medidas de gases CO_2 , atinge o valor máximo de 1,1066 mL no tempo de 47,5 horas.

Agradecimentos

O presente trabalho foi realizado com o apoio financeiro da CAPES/Embrapa. A Texas A&M University, Dept. Animal Science, pelo fornecimento dos dados e suporte, bem como a Embrapa Amazônia Oriental e a Embrapa Informática Agropecuária pelo grupo de par-

ceria científica da rede Pecu e AnimalChange.

Referências

ARAÚJO, D.R.C.A.; CARNEIRO, H.; SILVA, A.M.A; SILVA, U.L.; BERNARDI, T.M. Avaliação da degradabilidade *in vitro*, ácidos graxos voláteis e perdas energéticas através da produção total de gases, metano e dióxido de carbono de quatro espécies forrageiras em bovinos e ovinos. **Revista Científica de Produção Animal**, Salvador, v. 14, n. 2, p. 203-206, Dez. 2012.

BRUN, R.; REICHERT, P.; KUNSCH, H.R. Practical Identifiability analysis of large environmental Simulation Models. **Water resources research**, San Diego, v. 37, p. 1015-1030, 2001.

CABRAL, L.S.; VALADARES FILHO, S.C.; MALAFAIA, P.A.M.; LANA, R.P.; SILVA, J.F.C.; VIEIRA, R.A.M.; PEREIRA, E.S. Frações de carboidratos de alimentos volumosos e suas taxas de degradação estimadas pela técnica de produção de gases. **Revista Brasileira Zootecnia**, Viçosa, v. 29, n. 6, p. 2087-2098, Abr. 2000.

CALDERHEAD, B.; GIROLAMI, M. Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods. **Interface Focus**, London, v. 1, p. 821-835, Aug. 2011.

CAMPOS, F.P.; BOSE, M.L.V.; BOIN, C.; LANNA, D.P.D.; MORAIS, J.P.G. Avaliação do sistema de monitoramento computadorizado de digestão *in vitro*: 3. Desaparecimento da matéria seca e/ou FDN pela produção de gás. **Revista Brasileira Zootecnia**, Viçosa, v. 29, n. 2, p. 537-544, Abr. 2000.

CASTRO, G.H.F. GRAÇA, D.S.; GONÇALVES, L.C.; MAURICIO, R.M.; RODRIGUEZ, N.M.; BORGES, I.; TOMICH, .R. Cinética de degradação e fermentação ruminal da *Brachiaria brizantha* cv. Marandu colhida em diferentes idades ao corte. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 59, p. 1538-1544, Set. 2007.

DETMANN, E.; SILVA, J.F.C.; VÁSQUEZ, H. M.; HENRIQUES, L.T.; HADDADE, I.R. Cinética da degradação ruminal dos carboidratos de quatro gramíneas tropicais em diferentes idades de corte e doses de adubação nitrogenada: técnica de produção de gases. **Revista Brasileira de Zootecnia**, Viçosa, v. 38, n. 1, p. 149-158, Jan. 2009.

FARIAS, L.N; VASCONCELOS, V.R.; CARVALHO, F.F.R.; SARMENTO, J.L.R. Avaliação dos modelos logístico bicompartimental e de Gompertz na estimativa da dinâmica de fermentação ruminal *in vitro* do farelo e da torta de babaçu (*Orbignya martiana*). **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 63, n. 1, p. 136-142, Fev. 2011.

- FERNANDES, A.M.; QUEIROZ, A.C.; PEREIRA, J.C; LANA, R.P; BARBOSA, M.H.P.; FONSECA, D.M.; DETMANN, E.; CABRAL, L.C.; PEREIRA, E.S.; VITTORI, A. Fracionamento e cinética da degradação *in vitro* dos carboidratos constituintes da cana-de-açúcar com diferentes ciclos de produção em três idades de corte. **Revista Brasileira Zootecnia**, Viçosa, v. 32, n. 6, p. 1778-1785, Dez. 2003 .
- FLEGAL, J.M.; HUGHES, J. mcmcse: Monte Carlo Standard Errors for MCMC. R package version 1.0-1. Riverside, CA and Minneapolis, MN, 2012.
- FONDEVILA, M.; BARRIOS, A. The gas production technique and its application to the study of the nutritive value of forages. **Cuban Journal of Agricultural Science**, Havana, v. 35, p. 187-196, 2001.
- FORTALEZA, A.P.S.; SILVA, L.D.F.; BARBERO, R.P.; BARBOSA, M.A.A.F.; PRADO, M.R.; CALSAMIGLIA, S.B. Efeito da torta de nabo forrageiro sobre a cinética de fermentação e degradação ruminal *in vitro*. **Archivos de Zootecnia**, Córdoba, v. 62, n. 237, Mar. 2013.
- GELMAN, A.; VARLIN, J.B.; STERN, H.S.; RUBIN, D.B. **Bayesian Data Analysis**. 2 ed. Boca Raton: Chapman & Hall/CRC, FL, USA, 2004.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: BERNADO, J.M.; BERGER, J.O.; DAWID, A.P.; SMITH, A.F.M.(Ed.). **Bayesian Statistics 4**. New York: Oxford University Press, 1992. 526 p.
- GONÇALVES, A.L.; LANA, R.P.; RODRIGUES, M.T.; VIEIRA, R.A.M.; QUEIROZ, A.C.; HENRIQUE, D.S. Cinética de degradação de alguns volumosos usados na alimentação de cabras leiteiras por intermédio da técnica de produção de gases sob diferentes níveis de pH. **Revista brasileira Zootecnia**, Viçosa, v. 30, n. 6, p. 1004-1012, Abr. 2001.
- GROOT, J.C.J.; CONE, J.W.; WILLIAMS, B.A.; DEBERSAQUES, F.M.A.; LANTINGA, E.A. Multiphasic analysis of gas production kinetics for *in vitro* fermentation of ruminant feeds. **Animal Feed Science and Technology**, v. 64, p. 77-89, 1996.
- GUIMARÃES JR. R.; GONÇALVES, L.C.; MAURÍCIO, R.M.; PEREIRA, L.G.R.; TOMICH, T.R.; PIRES, D.A.A.; JAYME, D.G.; SOUSA, L.F. Cinética de fermentação ruminal de silagens de milheto. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte , v. 60, n. 5, p. 1174-1180, Set. 2008.
- HAARIO, H.; LAINE, M.; MIRA, A.; SAKSMAN, E. DRAM: Efficient Adaptive MCMC. **Statistics and Computing**, New York, v. 16, p. 339-354, 2006.
- JHA, M. Hydrologic Simulations of the Maquoketa River Watershed Using SWAT. **Working Paper**, Ames v. 9, n. 492, p. 1-24, Jun. 2009.
- LAINE, M. Adaptive MCMC methods with applications in environmental and geophysical models. **Finnish Meteorological Institute Contributions**, Helsinki, v. 69, p. 1-48, 2008.

LOAGUE, K.; GREEN, R.E. Statistical and graphical methods for evaluating solute transport models: Overview and application. **Journal of Contaminant Hydrology**, London, v. 7, n. 1-2, p. 51-73, Ago. 1991.

MAKKAR, H.P.S. Applications of the *in vitro* gas method in the evaluation of feed resources, and enhancement of nutritional value of tannin-rich tree/browse leaves and agro-industrial by-products. In: **Scientific Journals And Conference**, 1., 2002, Cairo. Proceedings. Cairo, Egypt: FAO/IAEA, division of nuclear Techniques in Food and Agriculture, 2002. p. 23-42.

MANLY, B.F.J. **Métodos estatísticos multivariados: uma introdução**. 3 ed. Porto Alegre: Brookman, 2008, 229 p.

MANTEL, N. The detection of disease clustering and a generalized regression approach. **Cancer Research**, London, v. 27, p. 209-220, 1967.

MARTÍNEZ-TERUEL, A.; MEGÍAS, M.D.; HERNÁNDEZ, F.; MADRID, J.; SALMERÓN, D.; CANO, J. A. Objective Bayesian vs. least squares estimation for by-products degradability with different rumen fluids. **Canadian Journal of Animal Science**, Ottawa, v. 89, p. 273-277, 2009.

MAURÍCIO, R.M.; PEREIRA, L.G.R.; GONCALVES, L.C.; RODRIGUEZ, N.M. Relação entre pressão e volume para implantação da técnica *in vitro* semi-automática de produção de gases na avaliação de forrageiras tropicais. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 55, n. 2, p. 216-219, Abr. 2003.

MAZUCHELI, J.; ACHCAR, J.A. Algumas considerações em regressão não-linear. **Acta Scientiarum**, Maringá, v. 24, n. 6, p. 1761-1770, 2002.

MELLO, R.; MAGALHÃES, A.L.R.; BREDAS, F.C.; REGAZZI, A.J. Modelos para ajuste da produção de gases em silagens de girassol e milho. **Pesquisa agropecuária brasileira**, Brasília, v. 43, n. 2, p. 261-269, Fev. 2008

MIZUBUTI, I.Y. RIBEIRO, E.L.A.; PEREIRA, E.S.; PINTO, A.P.; FRANCO, A.L.C.; SYPERRECK, M.A.; DÓREA, J.R.R.; CUNHA, G.E.; CAPELARI, M.G.M.; MUNIZ, E.B. Cinética de fermentação ruminal *in vitro* de alguns co-produtos gerados na cadeia produtiva do biodiesel pela técnica de produção de gás. **Ciências Agrárias**, Londrina, v. 32, n. 4, p. 2021-2028, Jul. 2011.

MODESTO, E.C.; SANTOS, G. T. S.; Vilela, D.; SILVA, D.C.; FAUSTINO, J.O.; JOBIM, C.C.; DETMANN, E.; ZAMBOM, M.A. ; MARQUES, J.A. Caracterização químico-bromatológica da silagem do terço superior da rama de mandioca, **Acta Scientiarum Animal Sciences**, Maringá, v. 26, n. 1, p. 137-146, 2004.

MORAIS, R.K.O.; SILVA, A.M.A.; BEZERRA, L.R.; CARNEIRO, H. MOREIRA, M.N.; MEDEIROS, F.F. *In vitro* degradation and total gas production of byproducts generated in the biodiesel production chain. **Acta Scientiarum Animal Sciences**, Maringá, v. 37, n. 2, p. 143-148, Jun. 2015.

MORÉ, J.J. The Levenberg-Marquardt algorithm: implementation and theory. In: WATSON G.A. (Ed.) **Numerical Analysis**, Berlin: Springer-Verlag, 1978. p. 105-116.

MOREIRA, P.C.; REIS, R.B.; REZENDE, P.L.P.; WASCHECK, R.C.; MENDONÇA, A.C.; DUTRA, A.R. Produção cumulativa de gases e parâmetros de France avaliados pela técnica semiautomática in vitro de fontes de carboidratos de ruminantes. **Revista Brasileira de Saúde e Produção Animal**, Salvador, v. 11, n. 2, p. 452-462, Dez. 2010.

MORGADO, E.S.; EZEQUIEL, J.M.B.; HOMEN JÚNIOR, A.C.; GALZERANO, L. Potencial de produção dióxido de carbono in vitro de dietas utilizadas em dietas para ovinos. Goiás, **Ciências Animal Brasileira**, v. 14, n. 4, p. 413-417, Dez. 2013.

NOGUEIRA, U.T.; MAURÍCIO, R.M.; GONÇALVES, L.C. Comparação de substratos com diferentes quantidades de carboidratos solúveis utilizando a técnica in vitro semi-automática de produção de gases, **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 58, n. 4, p. 633-641, Maio 2006.

NRC. **Nutrient Requirements of Beef Cattle**, 7th revised edn. National Academy Press, Washington, DC, 1996, 242 p.

PELL, A.N.; MOLINA, D.O.; SCHOFIELD, P. Measurement of gas production in vitro. In: DEAVILLE, E.; WILLIAMS, B.A.; CONE, J.; TAMMINGA, S.; GIVENS, D.I. (Ed.). **Gas production: fermentation kinetics for feed evaluation and to assess microbial activity**. Wageningen: BSAS, 2000. p. 1-12.

PEREIRA, E.M. et al. Determinação in vitro do potencial de produção de metano e dióxido de carbono de líquido ruminal proveniente de bovinos de diferentes categorias, **Archivos latinoamericanos de producción animal**, Maracaibo, v. 14, n. 4, p. 120-127, Jul. 2006.

PINHEIRO, J.; BATES, D.; DEBROY, S.; SARKAR, D.; R Core Team (2014). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117, Disponível em: <<http://CRAN.Rproject.org/package=nlme>>. Acesso em: 11 jun. 2015.

PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K. CODA: Convergence Diagnosis and Output Analysis for MCMC, **R News**, New Zealand, v. 6, n. 1, p. 7-11, Mar. 2006.

PRESS, W.H.; TEUKOLSKY, S.A.; VETTERLING, W.T.; FLANNERY B.P. **Numerical Recipes in FOR-TRAN. The Art of Scientific Computing**. 2 ed. Cambridge University Press, NY, USA, 1992, 554 p.

RAFTERY, A.L.; LEWIS, S. Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. **Statistical Science**, Hayward, v. 7, n. 4, p. 493-497, Nov. 1992.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2014. Disponível em: <<http://www.R-project.org>>. Acesso em: 20 mar. 2014.

- RODRIGUES, R.T.S. et al. Produção de gases e digestibilidade in vitro de silagens de maniçoba aditivadas com coproduto vitivinícola. **Revista Brasileira de Ciências Agrárias**, Recife, v. 7, n. 4, p. 684-690, Dez. 2012.
- SÁ, J.F.; PEDREIRA, M.S.; SILVA, F.F.; FIGUEIREDO, M.P.; REBOUÇAS, G.M.N.; SOUZA, D.R. Cinética da fermentação in vitro do capim-Marandu em diferentes idades de corte. **Acta Scientiarum Animal Sciences**, Maringá, v. 33, n. 3, p. 225-231, Dez. 2011.
- SCHOFIELD, P.; PITT, R.E.; PELL, A.N. Kinetics of fiber digestion from in vitro gas production. **Journal Animal Science**, Champaign, v. 72, p. 2080-2991, 1994.
- SILVA, A.R. biotools: Tools for Biometry and Applied Statistics in Agricultural Science. R package version 2.2., 2015. Disponível em : <<http://CRAN.R-project.org/package=biotools>>. Acesso em: 04 de novembro de 2015.
- SILVA, C.F.P.G.; FIGUEIREDO, M.P.; PEDREIRA, M.S.; BERNARDINO, F.S.; FARIAS, D.H.; AZÊVEDO, J.A.G. Cinética e parâmetros de fermentação ruminal in vitro de silagens de parte aérea e raízes de mandioca. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 64, n. 6, p.1639-1648, Jun. 2012.
- SNIFFEN, C.J.; O'CONNOR, J.D.; VAN SOEST, P.J.; FOX, D.G.; RUSSELL, J.B. A net carbohydrate and protein system for evaluating cattle diets. II. Carbohydrate and protein availability. **Journal of Animal Science**, Champaign, v. 70, n. 7, p. 3562-3577, Nov. 1992.
- SOETAERT, K.; PETZOLDT, T.; SETZER, R.W. Solving Differential Equations in R: Package deSolve. **Journal of Statistical Software**, Los Angeles, v. 33, n.9, p. 1-25, Dec. 2010.
- SOETAERT, K.; PETZOLDT, T. Inverse modelling, sensitivity and Monte Carlo analysis in R using package FME. **Journal of Statistical Software**, Los Angeles, v. 33, p. 1-28, 2010.
- STATISTICA for Windows. Release 6. Copyright Stat Soft. 1984-2001
- STEINFELD, H.; GERBER, P.; WASSENAAR, T.; CASTEL, V.; ROSALES, M.; HAAN, C. **Livestock's long shadow: Environmental issues and options**. Rome, Italy, 2006. 377 p.
- STEWART, J. **Cálculo**. São Paulo: Cengage Learning, 2011. v. 2, 1077 p.
- TEDESCHI, L.O. Assessment of the adequacy of mathematical models. **Agricultural Sciences**, Barking, v. 89, p. 225-247, 2006.
- TEDESCHI, L.O.; FOX, D.G.; KONONOFF, P.J. A dynamic model to predict fat and protein fluxes associated with body reserve changes in cattle. **Journal of Dairy Science**, New York, v. 96, p. 2448-2463, 2013.
- TEDESCHI, L.O.; KONONOFF, P.J.; KARGES, K.; GIBSON, M.L. Effects of chemical composition variation on the dynamics of ruminal fermentation and biological value of corn milling (co) products. **Journal of Dairy Science**, New York, v. 92, p. 401-413, 2009.

TEDESCHI, L.O.; SCHOFIELD, P.; PELL; A.N. Determining feed quality for ruminants using in vitro gas production technique. 1. Building an anaerobic fermentation chamber. In **The 4th Workshop on Modeling in Ruminant Nutrition: Application of the Gas Production Technique**, Juiz de Fora, MG, Brazil, 2008, p. 1-16.

TOMICH, T.R.; GONÇALVES, L.C.; MAURÍCIO, R.M.; PEREIRA, L.G.R.; RODRIGUES, J.A.S. Composição bromatológica e cinética de fermentação ruminal de híbridos de sorgo com capim-sudão. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, Belo Horizonte, v. 55, n. 6, p. 747-755, Set. 2003.

U.S. Environmental Protection Agency (USEPA). Global anthropogenic CO_2 greenhouse gas emissions: 1990-2020. Washington: USEPA, 2006.

U.S. Environmental Protection Agency (USEPA). Inventory of U.S. greenhouse gas emissions and sinks: 1990-2009. Washington: USEPA, 2011.

VAN SOEST, P.J. Development of a comprehensive system of feed analyses and its application to forages. **Journal of Animal Science**, Champaign, v. 2, n. 1, p. 119-128, 1967.

WANG, G.; CHEN, S. Evaluation of a soil greenhouse gas emission model based on Bayesian inference and MCMC: Parameter identifiability and equifinality. **Ecological Modelling**, Scotland, v. 255, p. 107-116, 2013.

WASCHECK, R.C.; REZENDE, P.L.P.; MOREIRA, P.C.; REIS, R.B.; ROSA, S.R.A.; A.C. MENDONÇA. Degradabilidade e produção de gases in vitro de fontes energéticas alternativas na alimentação de ruminantes. **Acta Scientiarum Animal Sciences**, Maringá, v. 32, n. 4, p. 425-430, Dez. 2010.

WU, Y; LIU, S. Automating calibration, sensitivity and uncertainty analysis of complex models using the R package Flexible Modeling Environment (FME): SWAT as an example. **Environmental Modelling & Software**, Scotland, v. 31, p. 99-109, 2012.

WU, L.; LIU, S. ; LIU, D.; FANG, Z.; XU, H. Modelling and forecasting CO_2 emission in the BRICS (Brazil, Russia, India, China, and South Africa) countries using a novel multi-variable grey model. **Energy**, Pequim, v. 79, n. 1, p. 489-495, 2015.

3 MODELO MATEMÁTICO DINÂMICO PARA ESTIMAR INDIVIDUALMENTE O DESEMPENHO DE BOVINOS DE CORTE CRUZADOS

Resumo

Não existem modelos dinâmicos de crescimento e composição química sendo usados em confinamentos comerciais capazes de gerar curvas de crescimento baseadas no indivíduo no Brasil. Este cenário está prestes a acontecer com os sistemas de monitoramento do desempenho individuais disponíveis. A estimação da composição química pelo Modelo de Crescimento de Davis (MCD), proposto por Oltjen et al. (1986) e a atualização em Oltjen et al. (2000), podem ajudar a determinar o ponto ótimo de abate de bovinos, permitindo que os produtores cumpram às especificações rigorosas do mercado internacional e aumentem a rentabilidade. O objetivo deste estudo foi apresentar uma abordagem para estimação dos parâmetros determinados individualmente pelo MCD para prever a dinâmica do crescimento e composição corporal de cada animal. Dados de grupos de bovinos, machos não castrados e fêmeas, com dois sistemas diferentes de produção (intensivo e extensivo) para cada parâmetro de ajuste foram usados para ajustar o parâmetro individual. Análises de covariâncias multivariadas foram realizadas baseadas nas estimativas individuais dos parâmetros ajustados pelo MCD. Vetores aleatórios iniciais com distribuição de probabilidade uniforme foram realizados para iniciação da função de otimização do sistema não-linear do MCD. A menor soma de quadrado residual para cada indivíduo foi determinada utilizando o método de otimização L-BFGS-B do *software* R. A otimização foi executada com vários pontos de partida $N = 100$ para evitar ótimos locais. Concluiu-se que a análise de covariância multivariada dos vetores de parâmetros individuais revelaram efeitos pronunciados quanto aos sistemas de produção, gêneros e a interação significativa entre esses fatores.

Palavras-chave: Multivariada; Ajuste individual; Otimização; Equações diferenciais; Não-linear

Abstract

Dynamic models to generate growth curves for individual growth curves for bovines are commercially unavailable in Brazil. This scenario is about to as individual performance monitoring systems became available. The prediction of chemical composition by Davis Growth Model (DGM), proposed by Oltjen et al. (1986) and the update in Oltjen et al. (2000), can help to determine the optimal point of slaughter of cattle allowing the producers to meet stringent specifications of the international market and to increase profitability. The aim of this study was to present an approach for prediction of individually determined parameter values of the DGM to estimate the dynamics of growth and body composition for each animal. Data from groups of male and female bovines in two different production systems (intensive and extensive) were used for individual parameter fitting. Covariance multivariate analyzes were performed based on the individual parameter estimates by DGM. Uniform random initial vectors were performed to initiation of the optimization function of the nonlinear system DGM. Least residual sum of squares for each individual was determine using the L-BFGS-B method of the *software* R. The optimization was run with multiple starting points $N=100$ to avoid local optima. Covariance multivariate analyzes of the individual parameter vectors revealed pronounced effects of production systems, genres and significant interaction between those factors.

Keywords: Multivariate; Individual fitting; Optimization; Differential Equations; Nonlinear

3.1 Introdução

Não existem modelos dinâmicos de crescimento e composição química sendo usados em confinamentos comerciais capazes de gerar curvas de crescimento baseadas no indivíduo. Este cenário está prestes a mudar, pois implantação de sistemas que monitore indivíduos e não mais lotes ou rebanhos começam a ser usados. Sistemas mecanicistas acoplados ao monitoramento de cada indivíduo poderão auxiliar nutricionistas e geneticistas encontrarem soluções que melhorem a produtividade e a lucratividade nas fazendas. Por outro lado, uma das grandes deficiências de parametrizar modelos capazes de prever as curvas para os indivíduos são os escassos dados de composição química corporal (i.e., gordura e proteína). Animais são abatidos antes de atingirem o peso da idade adulta, o que cria uma segunda deficiência, já que praticamente não há informação para os animais nesta fase.

Existe ainda a ineficiência de modelos matemáticos para representarem um sistema real de crescimento bovino. Entre os modelos existentes na literatura científica, os estáticos não estimam os estados transitórios do animal. Curvas de regressão não-linear dependentes do tempo usadas para descrever as trajetórias de peso dos animais e sua composição são também limitadas por não estimarem a forma como a energia é contida. Desta forma estes modelos negligenciam as possíveis interações no metabolismo e composição do ganho (exemplo: Bhowmick e Bhattacharya, Gompertz, Logístico, Von Bertalanffy, Weibull e Richards).

Os modelos matemáticos dinâmicos de crescimento de bovinos de corte são potencialmente mais eficientes por contemplarem variáveis que expressam o metabolismo e a disponibilidade de nutrientes. O Modelo de Crescimento de Davis (MCD) proposto por Oltjen et al. (1986) é baseado em conceitos de hiperplasia e hipertrofia do animal, que advêm da utilização de variáveis de estado como a massa de proteína, massa de DNA e massa de gordura que representam melhor o sistema de crescimento de bovinos. Essas características contempladas no modelo garantem vastas condições de aplicação que os modelos derivados empiricamente.

Embora já tenham sido publicados e analisados dados sobre exigências nutricionais e crescimento de bovinos no Brasil (Marcondes et al., 2010ab, b; Paulino et al., 2010; Sainz et al., 2006; Valente et al., 2013) estudos de parametrizações do MCD são escassos e nenhum deles incluem predições individuais e inferências estatísticas a partir desses resultados. Deste modo, o objetivo desse trabalho foi apresentar pela primeira vez uma abordagem para predições individuais dos animais para taxa de ganho de peso e energia de manutenção baseada na dinâmica de crescimento e composição química corporal do MCD e com comparação multivariada entre os grupos (gêneros e sistemas).

3.2 Material e Métodos

Os métodos foram divididos em três partes: base de dados, com a descrição do experimento; modelo, com os detalhes da descrição das variáveis e parâmetros do MCD; Estatística Computacional, em que são descritos os procedimentos da solução dos sistemas de Equações

Diferenciais Ordinárias (EDO), o cálculo da função objetivo, a otimização da função objetivo e as análises multivariadas oriundas das predições dos parâmetros individuais dos animais pelo MCD.

Base de dados. Todos os procedimentos foram realizados de acordo com as normas estabelecidas nas Estações de Pesquisa da Embrapa e do Comitê de ética da Escola Superior de Agricultura “Luiz de Queiroz”. O banco de dados do experimento foi conduzido em 2007 na Embrapa Gado de Corte. A base de dados utilizada neste estudo (confinamento em baias individuais após a recria) foi composta de 30 animais dos seguintes genótipos e gêneros: $1/2$ Red Angus $\times$ $1/4$ Caracu $\times$ $1/4$ Nelore [RCN, $n = 10$ fêmeas (F) e $n = 10$ machos (M)] e Red Angus $\times$ Nelore (RN, $n = 5$ F e $n = 5$ M). Esses animais vieram de sistemas distintos antes do período de desmame: a pasto (Sistema Extensivo, Sist. Extensivo) e confinado (Sistema Intensivo, Sist. Intensivo). No Sistema Extensivo pares de fêmeas e seus bezerros (pares vaca-bezerro) foram mantidos em pastagem com oferta de sal mineral (15,14% Na, 10,56% Ca, 8,92% P, 1,07% Mg, 6,06% S, 6,691 ppm Zn, 2,829 ppm Fe, 1,153 ppm Cu, 797 ppm Mn, 90 ppm I, 31 ppm S e 24 ppm Co, bases de matéria seca). Os pares vaca-bezerro do Sistema Extensivo não foram monitorados quanto ao consumo do pasto ou sal mineral na pré-desmama (período compreendido entre o nascimento e a desmama).

Aos bezerros do Sistema Intensivo na pré-desmama foram oferecidos uma única dieta simulando um pasto de boa qualidade [2,3 Mcal EM/kg de Matéria Seca (MS), 12,7% de Proteína Bruta (PB), base de MS] como vacas numa base *ad libitum* de 33 dias de idade. O consumo de DM de cada vaca foi ajustado individualmente em intervalos de 28 dias, minimizando as mudanças no Peso Corporal (PV_z) e mantendo os escores da condição corporal (ECC) constante, dieta e o manejo dos animais são apresentados em detalhes em (Albertini et al., 2012). Após o desmame e período de recria (225 ± 14 dias) os bezerros de ambos os sistemas foram alimentados individualmente com dieta mista contendo 2,84 Mcal EM/kg MS e 13,9% PB, em baias individuais. O banco de dados utilizado neste estudo se refere aos animais terminados em confinamento após a recria. Os animais foram abatidos quando atingiram 6 milímetros de gordura subcutânea nas 12 – 13^a costelas e o período máximo de confinamento foi de 147 dias.

No início do confinamento, a composição química corporal inicial foi obtida a partir do abate de 2 a 4 animais de cada genótipo. Além disso, no final de confinamento, os animais foram abatidos e energia retida foi estimada com base na composição da 9 – 10 – 11^a costelas.

Composição química corporal foi estimada a partir de regressões lineares com as porcentagens de água e gordura na 9 – 10 – 11^a costelas. A metodologia original de Hankins e Howe (1946) foi modificada, de modo que as seções das costelas inteiras (ossos e tecidos moles) foram triturados. A composição da 9 – 10 – 11^a costelas inteira foi utilizada. A proteína e as cinzas no corpo vazio foram calculados a partir da gordura e água estimada usando a razão 80:20 de proteína e de cinzas na gordura livre (Reid et al., 1955). A concentração de energia utilizada para a proteína e gordura foi 5,497 e 9,390 Mcal/kg de PV_z , respectivamente.

As variáveis utilizadas para parametrizar o modelo Oltjen et al. (1986) foram

baseadas no período experimental após o desmame: gênero, grupo genético, idade, sistema de produção, o peso corporal inicial (PV_i), kg ; Tempo de confinamento (TC), dias; ingestão de matéria seca (IMS), kg IMS/d ; peso corporal em jejum inicial (PV_{ji}), kg ; peso corporal em jejum final (PV_{jf}), kg ; peso do corpo vazio inicial (PV_{zi}), kg ; peso de corpo vazio final (PV_{zf}), kg ; peso corporal de proteína inicial ($PROT_{PV_{zi}}$), kg ; peso corporal de gordura inicial ($GORD_{PV_{zi}}$), kg ; o peso corporal de proteína final ($PROT_{PV_{zf}}$), kg ; o peso corporal de gordura final ($GORD_{PV_{zf}}$), kg . O conjunto de dados agregadas por gênero e sistema foram resumidos nas Tabelas 3.1 e 3.2.

Tabela 3.1 – Resumo com as análise descritiva [médias, desvio padrão (DP), mínimo (Min) e máximo (Max)] das variáveis de entrada do modelo - Fêmeas

Sistema	Mensurações referentes ao período pós-desmame				
pré-desmame	Variáveis	Média	DP	Min	Max
Sistema Extensivo	PV_i, kg	293,7500	25,3295	267,0000	328,0000
	TC, dia	90,0000	50,9509	47,0000	147,0000
	IMS, kg IMS/dia	7,4000	1,1234	6,3500	8,9900
	PV_{ji}, kg	293,7500	25,3295	267,0000	328,0000
	PV_{jf}, kg	386,0000	27,6043	347,0000	412,0000
	PV_{zi}, kg	259,7500	22,3960	236,0000	290,0000
	PV_{zf}, kg	348,5000	23,7276	315,0000	371,0000
	$PROT_{PV_{zi}}, kg$	44,4750	3,8621	40,4000	49,7000
	$GORD_{PV_{zi}}, kg$	46,3000	3,9874	42,1000	51,7000
	$PROT_{PV_{zf}}, kg$	58,0500	3,6990	53,0000	61,9000
	$GORD_{PV_{zf}}, kg$	74,4000	7,3362	63,4000	78,3000
Sistema Intensivo	PV_i, kg	282,2222	24,4580	236,0000	308,0000
	TC, dia	70,1111	22,5745	47,0000	119,0000
	IMS, kg IMS/dia	7,6944	0,8675	6,2100	9,1500
	PV_{ji}, kg	282,2222	24,4580	236,0000	308,0000
	PV_{jf}, kg	368,3333	25,5587	323,0000	413,0000
	PV_{zi}, kg	249,3333	21,5986	209,0000	272,0000
	PV_{zf}, kg	328,7778	25,8397	286,0000	370,0000
	$PROT_{PV_{zi}}, kg$	42,5111	3,7468	35,7000	46,6000
	$GORD_{PV_{zi}}, kg$	46,1111	3,9228	37,2000	49,9000
	$PROT_{PV_{zf}}, kg$	54,8779	4,7023	47,4000	62,0000
	$GORD_{PV_{zf}}, kg$	69,3779	4,7874	62,6000	75,8000

Modelo. Os modelos de simulação dinâmica têm sido desenvolvidos para acompanhar precisamente as trajetórias de ganho de peso e composição corporal durante o crescimento de animais, particularmente de bovinos. Oltjen et al. (1986) propuseram o uso de uma forma mais geral, com base no número de células e mecanismos de tamanho de crescimento, para prever a síntese líquida de proteína e integrou o modelo de energia líquida para estimar o ganho de gordura e tecido magro. O modelo foi atualizado por Oltjen et al. (2000) e é composto por três variáveis de estado correspondente a massa de DNA (DNA), massa de proteína ($PROT$) e massa de gordura ($GORD$) totais, descrito no seguinte sistema de equações em (3.1):

Tabela 3.2 – Resumo com as análise descritiva [médias, desvio padrão (DP), mínimo (Min) e máximo (Max)] das variáveis de entrada do modelo - Machos

Sistema	Mensurações referentes ao período pós-desmame				
pré-desmame	Variáveis	Mean	DP	Min	Max
Sistema Extensivo	PV_i, kg	320,8000	44,9966	268,0000	388,0000
	TC, dia	132,6000	32,1993	75,0000	147,0000
	$IMS, kg\ IMS/dia$	8,68600	1,1030	7,6700	10,1400
	PV_{ji}, kg	320,8000	44,9966	268,0000	388,0000
	PV_{jf}, kg	483,6000	52,6573	428,0000	568,0000
	PV_{zi}, kg	283,8000	39,7077	237,0000	343,0000
	PV_{zf}, kg	436,2000	51,3245	378,0000	518,0000
	$PROT_{PV_{zi}}, kg$	50,1000	7,0228	41,9000	60,6000
	$GORD_{PV_{zi}}, kg$	38,9200	5,4733	32,5000	47,1000
	$PROT_{PV_{zf}}, kg$	75,1200	8,6033	64,3000	88,0000
	$GORD_{PV_{zf}}, kg$	75,1000	12,9591	67,5000	98,1000
Sistema Intensivo	PV_i, kg	329,0000	30,7530	270,0000	366,0000
	TP, dia	108,5556	31,5084	75,0000	147,0000
	$IMS, kg\ IMS/dia$	8,8078	0,8024	7,5200	9,7800
	PV_{ji}, kg	329,0000	30,7530	270,0000	366,0000
	PV_{jf}, kg	492,5556	28,1119	445,0000	544,0000
	PV_{zi}, kg	290,6667	27,0231	239,0000	323,0000
	PV_{zf}, kg	442,4444	27,8303	395,0000	488,0000
	$PROT_{PV_{zi}}, kg$	51,0111	4,7166	42,2000	57,2000
	$GORD_{PV_{zi}}, kg$	42,8555	5,6736	32,8000	51,0000
	$PROT_{PV_{zf}}, kg$	75,8111	4,6934	67,9000	83,9000
	$GORD_{PV_{zf}}, kg$	78,9000	6,1650	69,0000	85,7000

$$\left\{ \begin{array}{l} \frac{\partial DNA}{\partial t}(g/dia) = k_1 NUT_1 (DNA_{max} - DNA) \\ \frac{\partial PROT}{\partial t}(kg/dia) = NUT_2 k_2 DNA^{0,73} - k_3 PROT^{0,73} \\ \frac{\partial GORD}{\partial t}(kg/dia) = \frac{\left(\frac{IEM - EM_{MAINT}}{EL_m} \right) EL_g \frac{\partial PROT}{\partial t} E_{Prot}}{EGord}, \end{array} \right. \quad (3.1)$$

em que k_1 é a taxa de acréscimo de DNA, k_2 é a taxa de síntese de proteínas, k_3 é a taxa de degradação de proteína, DNA_{max} é a quantidade de DNA máximo em todo o corpo; NUT_1 e NUT_2 são os efeitos de ingestão de energia (Oltjen et al., 1986) e elas são representadas nas equações 3.2 e 3.3:

$$NUT_1 = -0,70 + 1,70P \quad (3.2)$$

$$NUT_2 = 0,83 + \frac{0,20P}{0,15 + P}, \quad (3.3)$$

em que

$$P = \frac{IEM}{IEM_{norm}} \quad (3.4)$$

é a razão entre a ingestão de energia metabolizável observada, IEM , e a ingestão de energia metabolizável de referência para alcançar um crescimento normal do animal, IEM_{norm} (MJ/dia) :

$$IEM_{norm} = (1,8210 - 0,00145(PV_z)) \times PV_z^{0,73}. \quad (3.5)$$

A taxa de deposição de gordura é calculada pelo quociente entre a quantidade de energia residual e o valor médio da energia de lípidos, de acordo com terceira equação do sistema (3.1). Em que FAT é a massa de gordura; $EProt$ e $EGord$ são as concentrações de energia de proteína e gordura, respectivamente; EL_m e EL_g representam a concentração de energia alimentar para manutenção e ganho, respectivamente; EM_{MAINT} é a exigência de energia metabolizável para manutenção

$$EM_{MAINT}(Mcal/dia) = \alpha(PV_z)^{0,75} \quad (3.6)$$

em que α é o parâmetro para ser estimado e representa a energia de manutenção requerida dos animais e PV_z pode ser estimado como

$$PV_z = GORD + \frac{PROT}{0,2201}. \quad (3.7)$$

Os parâmetros de interesse no modelo do Oltjen são k_1 , k_2 , k_3 , DNA_{max} e α (referentes ao sistema 3.1). A interpretação prática desses parâmetros é que refletem o genótipo dos animais. No entanto, k_2 e k_3 , possuem alta correlação, logo é inviável ajustar os dois parâmetros simultaneamente. Assim, uma alternativa citada na literatura científica, Sainz et al. (2006), Dantas (2007) e Oltjen et al. (1986), é manter o valor de k_2 constante em 0,0461. Para esse estudo, a proposta foi predizer os parâmetros, k_1 e α , para cada indivíduo.

Estatística Computacional. Para a solução numérica de problemas de valores iniciais de sistemas de primeira ordem de EDO, foi usado o pacote *deSolve* e a função *ode()* pelo método integrador *lsoda* do *software* R (Development Core Team, versão 3.1.0, 2014). O erro absoluto e relativo de tolerância foi tomado como 10^{-6} .

Dada uma solução do modelo e os dados observados, a função objetivo que mini-

miza o erro foi calculada pela seguinte equação

$$Res_i = \sum_{j=1}^p (\hat{y}_{ij} - y_{ij})^2 \quad (3.8)$$

em que Res_i representa o valor residual do i -ésimo indivíduo; $\hat{y}_{ij}$ e y_{ij} são o valores preditos e observados do MCD, respectivamente, do i -ésimo indivíduo na j -ésima variável do modelo, $i = 1, 2 \dots n$; $j = 1 \dots p$; n representando o número de animais e p o número de variáveis respostas, compreendidas pela massa de proteína final e massa de gordura final. Note que nenhum peso (padronização) foi considerada na função que minimiza a função residual individual, pois, cada animal era composto por duas mensurações apenas, uma para a massa de proteína e outra para a gordura.

O método L-BFGS-B foi usado para predição dos parâmetros. A convergência ocorre quando o fator de tolerância atinge 10^{-8} , correspondente aos valores da função objetivo entre uma iteração e outra. Para evitar problemas do algoritmo recair numa solução de mínimo local e não no mínimo global da soma de quadrados residual, uma função foi criada para gerar N vetores aleatórios iniciais, $\mathbf{p}_0 = [(k_1)_0, (\alpha)_0]$, de uma distribuição uniforme independente. Esses vetores foram passados para a entrada da função de otimização como valores iniciais para iniciar o algoritmo. Assim, para cada animal foi computado um total de N vetores preditos e N soma residuais correspondentes a cada um dos valores iniciais de vetores aleatórios realizados. O vetor de predições, $\mathbf{p}_m = [k_1, \alpha]$, que obteve o menor valor residual foi escolhido para compor a predição do animal. Para todos os grupos de tratamentos dos dados experimentais foi considerado $N = 100$.

Foram realizadas análises de variância univariadas e multivariadas para verificação de diferenças significativas no crescimento dos animais, baseados nos vetores preditos $\mathbf{p}_m$ pelo modelo MCD. Os tratamentos do experimento consistiu em todas as combinações entre os dois sistema (Extensivo e Intensivo) e os gêneros (fêmeas e machos), compondo no total de quatro tratamentos. Os genótipos dos animais eram homogêneos e consistiam em animais cruzados. O modelo estatístico utilizado na análise multivariada foi:

$$\mathbf{y}_{ijkp} = \boldsymbol{\mu}_p + \boldsymbol{\alpha}_{ip} + \boldsymbol{\beta}_{jp} + \boldsymbol{\gamma}_{ijp} + \boldsymbol{\rho}_{ijkp} + \boldsymbol{\epsilon}_{ijkp}, \quad (3.9)$$

em que $i = 1, \dots, s$, sendo i o índice referente aos sistemas (extensivos e intensivo); $j = 1, \dots, g$, sendo j o índice referente aos gêneros (fêmeas e machos não castrados); $k = 1, \dots, r_{(t)}$, sendo $r_{(t)}$ as repetições de cada tratamento t ($r_{(fêmeas\ sistema\ extensivo)} = 4$, $r_{(fêmeas\ sistema\ intensivo)} = 9$, $r_{(machos\ sistema\ extensivo)} = 5$ e $r_{(machos\ sistema\ intensivo)} = 9$, composto por um desbalanceamento com total de 27 unidades experimentais); $p = 1, \dots, v$, sendo p o indexador da variável resposta (taxa de acréscimo de DNA e energia de manutenção). $\mathbf{y}_{ijkp}$ é um vetor multivariado de observações da p -ésima variável sob efeito do i -ésimo sistema, j -ésimo gênero e k -ésima repetição; $\boldsymbol{\mu}_p$ é um vetor de médias geral inerente a todas as observações na p -ésima variável; $\boldsymbol{\alpha}_{ip}$ é um vetor de

média de efeitos do i -ésimo sistema na p -ésima variável; β_{jp} é um vetor de média de efeitos do j -ésimo gênero na p -ésima variável; γ_{ijp} é o vetor multivariado dos efeitos de interações entre o i -ésimo sistema e o j -ésimo gênero do rebanho, na p -ésima variável; ρ_{ijkp} é a covariável (vetor multivariado do efeito dos pesos iniciais associados às observações y_{ijkp}); ϵ_{ijkp} é o vetor multivariado de erros associados às observações y_{ijkp} .

O vetor de erros segue uma distribuição normal multivariada ($\epsilon_{ijkp} \sim N_2(\mathbf{0}, \Sigma)$). Os vetores correspondentes as diferentes unidades experimentais são distribuídos de forma independente. As hipóteses nulas de igualdade dos vetores de efeitos do sistema, dos efeitos de gênero e da interação entre esses dois efeitos que caracterizam a amostra são apresentadas, respectivamente, por

$$H_{0_i} : \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \vdots \\ \alpha_{1p} \end{bmatrix} = \begin{bmatrix} \alpha_{21} \\ \alpha_{22} \\ \vdots \\ \alpha_{2p} \end{bmatrix} = \dots = \begin{bmatrix} \alpha_{s1} \\ \alpha_{s2} \\ \vdots \\ \alpha_{sp} \end{bmatrix}, \quad (3.10)$$

$$H_{0_j} : \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \vdots \\ \beta_{1p} \end{bmatrix} = \begin{bmatrix} \beta_{21} \\ \beta_{22} \\ \vdots \\ \beta_{2p} \end{bmatrix} = \dots = \begin{bmatrix} \beta_{g1} \\ \beta_{g2} \\ \vdots \\ \beta_{gp} \end{bmatrix}, \quad (3.11)$$

$$H_{0_{ij}} : \begin{bmatrix} \gamma_{111} \\ \gamma_{112} \\ \vdots \\ \gamma_{1jp} \end{bmatrix} = \begin{bmatrix} \gamma_{121} \\ \gamma_{122} \\ \vdots \\ \gamma_{1jp} \end{bmatrix} = \dots = \begin{bmatrix} \gamma_{s11} \\ \gamma_{s12} \\ \vdots \\ \gamma_{sjp} \end{bmatrix}, \quad (3.12)$$

sendo α_{ip} a observação média do efeito do i -ésimo sistema na p -ésima variável; β_{jp} a observação média do efeito do j -ésimo gênero na p -ésima variável; γ_{ijp} a observação média do efeito das interações entre o i -ésimo sistema e o j -ésimo gênero do rebanho, na p -ésima variável. Todas as hipóteses foram testadas contra a hipótese alternativa (H_1), no qual, pelo menos um efeito diferente entre estes vetores podem ser observados. Para as análise de variância univariadas, um modelo similar ao caso multivariado foi utilizado, porém, não se tratando mais de vetores de médias. O modelo é descrito a seguir

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \rho_{ijk} + \epsilon_{ijk}. \quad (3.13)$$

em que $i = 1, \dots, s$, sendo i o índice referente aos sistemas (extensivos e intensivo); $j = 1, \dots, g$, sendo j o índice referente aos gêneros (fêmeas e machos não castrados); $k = 1, \dots, r_{(t)}$, sendo $r_{(t)}$ as repetições de cada tratamento t ($r_{(\text{fêmeas sistema extensivo})} = 4$, $r_{(\text{fêmeas sistema intensivo})} = 9$, $r_{(\text{machos sistema extensivo})} = 5$ e $r_{(\text{machos sistema intensivo})} = 9$, composto por um desbalanceamento com total de 27 unidades experimentais); y_{ijk} é o valor observado para o i -ésimo sistema, j -ésimo gênero e k -ésima repetição; μ é a médias geral inerente a todas as observações; α_i é o efeitos do i -ésimo

sistema; β_j é o efeitos do j -ésimo gênero; γ_{ij} é o efeito da interação entre o i -ésimo sistema e o j -ésimo gênero do rebanho; ρ_{ijk} é a covariável (efeito dos pesos iniciais associados às observações y_{ijk}); ϵ_{ijk} é o erro associado às observações y_{ijk} , $\epsilon_{ijk} \sim N(0, \sigma^2)$.

A análise de resíduos foi testada, usando testes de homogeneidade e normalidade, além da inspeção gráfica. Elipses gráficas foram traçadas para representar a região de confiança do centróide de cada grupo para verificar diferença entre os tratamentos.

3.3 Resultados e Discussões

Com base nas estimativas individuais obtidas para k_1 e α por meio da amostra original foram conduzidas análises de variância univariada e multivariada, quando fixadas as demais constantes (k_2 , k_3 e DNA_{max}), pelo método de L-BFGS-B. O fenômeno biológico explicado pelos parâmetros k_1 e α é que eles refletem as diferenças quanto à taxa de ganho de peso e quanto à exigência de energia para manutenção, respectivamente, em cada animal.

Foram averiguados os pressupostos da análise de variância, verificando a análise residual para os modelos de covariância univariada e multivariada, apresentados na seção de Estatística Computacional. Verificou-se a presença de três pontos discrepantes. Foram removidos os pontos discrepantes e para variável resposta k_1 foi aplicada uma transformação de Box-Cox (1964) que foi elevada uma uma potência ótima com $\lambda = -0,6000$. Assim, os resíduos passaram a apresentar distribuição normal e homogeneidade de variância.

Em uma análise exploratória para identificar possíveis covariáveis no modelo estatístico suscetível de explicar as variáveis respostas foram testados o peso inicial dos animais após a recria e a idade do animal no início do experimento. Assim foi realizado o gráfico de dispersão com base na idade inicial versus a variável resposta $k_1^{-0,6}$, Figura 3.1 (a); e a idade inicial versus a variável resposta α , Figura 3.1 (b). Não verificou-se um comportamento claro de relações funcionais entre as variáveis respostas e a idade.

Na Figura 3.2 é representado o gráfico de dispersão para identificar a existência de relações funcionais entre o comportamento do peso do animal no início do confinamento e as variáveis $k_1^{-0,6}$ (a) e α (b). Notou-se, um comportamento linear para ambas as variáveis respostas.

Para explorar o efeito da interação dos fatores foram construídas as Figuras 3.3 e 3.4 com base na média de cada nível dos fatores. Na ausência de interação entre os fatores é esperado retas paralelas, significando que o efeito de um fator é o mesmo, independente do nível do outro fator. A interação é positiva quando os fatores apresentam efeitos sinérgico. O efeito conjunto dos fatores, nesse caso, é aumentar o resultado da combinação de seus níveis mais altos. Quando os fatores têm efeitos antagônicos, a interação é negativa. Tanto a Figura 3.3 como a Figura 3.4 indicam existir interação, devendo explorar os estudos dos efeitos de um fator em cada nível do outro. Será apresentada a análise de variância para comprovação de interações e demais procedimentos da análise.

Figura 3.1 – Análise de covariância - dispersão das idades (dias) em relação aos parâmetros k_1 (a) e α (b)

Figura 3.3 – Interação entre gênero e sistema para a variável k_1

Figura 3.2 – Análise de covariância - dispersão do peso inicial (kg) em relação aos parâmetros k_1 (a) e α (b)

Figura 3.4 – Interação entre gênero e sistema para a variável α

Na Tabela 3.3 são apresentados os resultados da análise de covariância univariada considerando o parâmetro de estudo k_1 . A covariável peso inicial dos animais foi significativa. A interação foi significativa ao nível de 0,05 de probabilidade e portanto, foram realizados os desdobramentos dos fatores em cada nível do outro fator. Na Tabela 3.4 são apresentados o ganho da taxa de peso (kg/dia) médio e teste de Tukey ao nível de 0,05 de probabilidade. Não houve diferença significativa quanto a mudança do nível intensivo para o nível extensivo quanto ao fator gênero para o ganho de peso (k_1). Resultado esperado, uma vez que animais do sistema extensivo possuem uma taxa de crescimento semelhante durante o ganho compensatório em relação animais do sistema intensivo.

Tabela 3.3 – Análise de covariância univariada considerando a taxa de ganho de peso (k_1)

FV	GL	SQ	QM	F_{cal}	Pr>F
Gênero	1	269,4925	269,4925	66,05	< 0,0001**
sistema	1	1,4890	1,4890	0,36	0,5519
Gênero × sistema	1	18,8820	18,8820	4,63	0,0427
Peso inicial	1	60,5658	60,5658	14,84	0,0009**
Resíduo	22	89,7579	4,0799	—	
Total	26	23,4932	—	—	

Vários autores (Felício, 1981,1982; Sainz, 1996; Luchiari Filho e Moura, 1998) destacam o efeito do gênero para a qualidade da carcaça. Os machos não castrados crescem mais rapidamente e depositam menos gordura que as fêmeas, já os machos castrados exibem características intermediárias, Haddad et al. (2000).

Se animais dos dois tipos de gêneros (machos não castrados e fêmeas) fossem abatidos na mesma idade, o macho não castrado produziria uma carcaça mais pesada do que as fêmeas. Apesar dessas diferenças, o conteúdo de gordura será menor no caso do macho não castrado, seguido pela fêmea que teria a maior proporção de gordura (Haddad et al., 2000). Assim, do ponto de vista de precocidade de acabamento, os machos não castrados são considerados mais tardios do que as novilhas. Os resultados desse estudo consolidam a diferença entre fêmeas e machos não castrados com clima tropical e genótipos advindos de cruzamento sobre óptica do parâmetro das taxas de acréscimo de proteína.

A gordura de cobertura vem se tornando um importante indicador de qualidade, porque afeta diretamente a velocidade de resfriamento da carcaça, comportando-se como um isolante térmico e interferindo no processo de conversão do músculo em carne. Além disso, carcaças com melhor acabamento tendem a apresentar maior teor de gordura entremeada, que é depositada com maior intensidade na fase em engorda dos bovinos, após o término da fase de crescimento, Haddad et al. (2000). Deste modo, produtores rurais que lidam com gado de corte, podem ser beneficiados quanto as decisões no tempo de abate de cada animal individual, tomando o conhecimento da capacidade do seu gado para atender especificações do mercado e aumentando a rentabilidade. Do outro lado, o frigorífico, que pode melhorar a estrutura de pagamento ao produtor e a sua margem econômica. O frigorífico com informações dos animais com base no ponto ótimo econômico (POA, isto é, momento mais lucrativo do animal) pode atender seus clientes (mercado doméstico ou importador) antecipando informações (quantitativas, qualitativas) dos animais, padronizando-os e agregando valor à negociação. A indústria também pode melhorar os processos de logística na compra dos animais, na alocação da matéria prima dentro do frigorífico, bem como aperfeiçoar a distribuição da carne aos seus clientes. Estes processos no frigorífico podem ser otimizados maximizando a lucratividade da atividade.

Na Tabela 3.5 é apresentada a análise de covariância univariada considerando o parâmetro de estudo α , representando a energia para manutenção do animal. A interação dos fatores gêneros e sistema (relativa a pré-desmana) foi significativa e os resultados do teste de Tukey, ao

Tabela 3.4 – Comparação das médias de taxas de acréscimo de proteína (fêmeas sistema extensivo, $n = 4$; fêmeas sistema intensivo, $n = 9$; machos sistema extensivo, $n = 5$; machos sistema intensivo, $n = 9$) e teste de Tukey

	parâmetro k_1	
	Fêmeas	Machos
Intensivo	0,0191 a A	0,0090 a B
Extensivo	0,0237 a A	0,0073 a B

As médias seguidas da mesma letra minúscula nas colunas não diferem entre si, pelo teste de Tukey, ao nível de 0,05 de probabilidade. As médias seguidas da mesma letra maiúsculas nas linhas não diferem entre si, pelo teste de Tukey, ao nível de 0,05 de probabilidade.

nível de 0,05 de probabilidade é apresentada na Tabela 3.6.

Tabela 3.5 – Análise de covariância univariada considerando à exigência de energia para manutenção (α)

FV	GL	SQ	QM	F_{cal}	Pr>F
Gênero	1	0,0008	0,0008	13,6600	0,0013**
Sistema	1	2×10^{-6}	2×10^{-6}	0,0400	0,8528
Gênero $\times$ Sistema	1	0,0011	0,0011	17,6900	0,0004**
Peso Inicial	1	6×10^{-5}	6×10^{-5}	0,8500	0,3666
Resíduo	22	0,0013	6×10^{-5}	—	
Total	26	0,0034	—	—	

Tabela 3.6 – Comparação de médias de energia para manutenção (fêmeas sistema extensivo, $n = 4$; fêmeas sistema intensivo, $n = 9$; machos sistema extensivo, $n = 5$; machos sistema intensivo, $n = 9$) e Teste de Tukey

	parâmetro α	
	Fêmeas	Machos
Intensivo	0,1205 a A	0,1207 a B
Extensivo	0,1060 b A	0,1340 b B

As médias seguidas da mesma letra minúscula nas colunas não diferem entre si, pelo teste de Tukey, ao nível de 0,05 de probabilidade. As médias seguidas da mesma letra maiúsculas nas linhas não diferem entre si, pelo teste de Tukey, ao nível de 0,05 de probabilidade.

Animais que apresentam ganho de peso mais elevados antes de ser confinados podem ter ganho menor e pior conversão alimentar quando confinados, comparados com animais vindos de sistemas com alguma restrição alimentar (NRC, 2000). Nesse estudo, os resultados apontaram que o sistema referente a pré-desmama afetou a fase de terminação dos animais, comparando o parâmetro.

Houve diferença significativa do sistema intensivo quando comparada a manutenção média entre fêmeas e machos (Tabela 3.6). A diferença da fêmea para o macho no sistema intensivo, tem um aumento médio na energia para manutenção de 0,0002 unidades. A diferença da fêmea para o macho no sistema extensivo, tem um aumento médio na energia para manutenção de 0,0280 unidades. Os resultados foram esperados pois animais machos são mais pesados portanto requerem mais energia para manutenção que as fêmeas. Os resultados apontam também diferenças

significativas entre animais machos do sistema extensivo e machos intensivos. Sendo a maior média de energia de manutenção para machos extensivos. Esse comportamento nos animais machos era esperado pois os animais intensivos foram abatidos num período de menor tempo durante o confinamento em relação aos animais do sistema extensivo.

O período de tempo que os animais permanecem confinados num sistema de confinamento é um fator muito importante, pois a medida que o peso adulto atinge uma assíntota, passa a ocorrer maior deposição de gordura em detrimento da proteína, e a deposição de gordura promove menor ganho de peso (NRC, 2000; Santos e Balsalobre, 2000). Portanto, é importante para ao produtor de gado de corte tomar decisões quanto ao momento certo da venda desses animais respeitando a deposição de gordura.

Na Tabela 3.7 são apresentados os contrastes entre vetores de médias da análise de covariância multivariada bivariada considerando as variáveis respostas k_1 e α . Os dados apresentaram distribuição normal multivariada e homogeneidade de variância, como a transformação da variável k_1 , quando elevada a potência ótima $\lambda = -0,6000$ (Box-Cox, 1964). E os resultados apontam diferenças significativa entre os vetores de médias apresentados na hipótese da equação (3.12). Além disso, diferenças entre os vetores de médias dos contrastes comparados foram:

1. Fêmeas no sistema intensivo e fêmeas no sistema extensivo $\times$ machos no sistema intensivo e machos no sistema extensivo;
2. Fêmeas no sistema intensivo $\times$ fêmeas no sistema extensivo;
3. Machos no sistema intensivo $\times$ machos no sistema extensivo;
4. Fêmeas no sistema extensivo $\times$ machos no sistema intensivo;
5. Fêmeas no sistema intensivo $\times$ machos no sistema extensivo;
6. Fêmeas no sistema intensivo $\times$ machos no sistema intensivo;
7. Fêmeas no sistema extensivo $\times$ machos no sistema extensivo.

Na Figura 3.5 são apresentadas as regiões de confiança de 95% dos vetores médios para os quatro tratamentos, resultantes das combinações dos níveis de cada fator. Essas elipses permitem uma boa visualização de discriminação entre os tratamentos. Observa-se que o comportamento dos tratamentos apresentam diferença entre os vetores médios devido a não sobreposição entre eles, o que é confirmado pelo fato da interação (gênero $\times$ sistema) apresentar diferença nos níveis dos fatores nos estudos realizados para os casos univariados e por existir diferença entre os contrastes realizados na análise de variância multivariada.

Na Tabela 3.8 são apresentadas as distâncias quadráticas de *Mahalanobis* e valores descritivos de probabilidade pelo teste F para os contrastes multivariados, considerando as combinações entre os níveis do gênero e os níveis do sistema. A maior distância entre os vetores médios dos grupos foram entre as fêmeas extensivas e os machos extensivos. E a menor diferença

Tabela 3.7 – Análise de covariância multivariada considerando os parâmetros k_1 e α

Contrastes	Wilks	F	Num GL	Den Gl	Pr>F
(Tratamentos)	0,1234	12,9200	(6)	42	<0,0001**
FSI ¹ FSE ² vs. MSI ³ MSE ⁴	0,2370	49,6200	2	21	<0,0001**
FSI vs. FSE	0,6392	6,1600	2	21	<0,0078**
MSI vs. MSE	0,5721	8,4300	2	21	<0,0021**
FSE vs. MSI	0,2941	25,2000	2	21	<0,0001**
FSI vs. MSE	0,2440	32,5300	2	21	<0,0001**
FSI vs. MSI	0,4195	14,5300	2	21	<0,0001**
FSE vs. MSE	0,1810	47,4900	2	21	<0,0001**
Peso inicial	0,6870	7,0900	2	21	<0,0044**

¹fêmeas no sistema intensivo; ²fêmeas no sistema extensivo; ³machos no sistema intensivo; ⁴ machos no sistema extensivo.

evidenciado pelos vetores médios entre fêmeas e machos no sistema intensivo. Salienta-se que os grupos apresentaram normalidade multivariada e homogeneidade de variância.

Tabela 3.8 – Distâncias quadráticas de *Mahalanobis* e valores descritivos de probabilidade pelo teste F para os contrastes multivariados, considerando as combinações entre os níveis do gênero e os níveis do sistema

Gênero/Sistema		Tratamentos			
		FSI	FSE	MSI	MSE
Tratamentos	FSI ¹	0,0000	6,7014	5,7732	22,5776
	FSE ²	0,0053	0,0000	17,6837	38,0871
	MSI ³	0,0096	<0,0001	0,0000	8,4947
	MSE ⁴	<0,0001	<0,0001	0,0018	0,0000

Os valores de probabilidade dos contrastes pelo teste F estão abaixo da linha diagonal, ao valor nominal de 5% de significância (valor $p < 0,05$). As distâncias quadráticas de *Mahalanobis* estão acima da linha diagonal; ¹fêmeas no sistema intensivo; ²fêmeas no sistema extensivo; ³machos no sistema intensivo; ⁴ machos no sistema extensivo.

Figura 3.5 – Elipses de 95% de confiança dos vetores de médias para as fêmeas no sistema intensivo, fêmeas no sistema extensivo, machos no sistema intensivo e machos no sistema extensivo de acordo com as variáveis k_1 transformada [potência ótima $\lambda = -0,6000$ (Box-Cox, 1964)] e α

3.4 Conclusões

Os vetores médios de efeitos quanto à energia de manutenção e o acréscimo de proteína parametrizados por meio de modelagem dinâmica do modelo de Crescimento de Davis de forma individual possuem efeitos pronunciados quanto as interações entre sistemas e gênero (valor $p < 0,01$), ou seja, efeitos ambientais com interação dos hormônios sexuais da diferença entre os gêneros apresentam diferença entre a relação a taxa média de acréscimo de proteína e a média de energia de manutenção.

Agradecimentos

O presente trabalho foi realizado com o apoio financeiro da CAPES/Embrapa. A Embrapa Gado de Corte, ao Laboratório Crescimento de Nutrição Animal ESALQ/USP pelo fornecimento dos dados e dos subsídios, bem como a Embrapa Amazônia Oriental e a Embrapa Informática Agropecuária pelo grupo de parceria científica da rede Pecuária e AnimalChange.

Referências

ALBERTINI, T.Z.; MEDEIROS, S.R.; TORRES JÚNIOR, R.A.A. ZOCCHI, S.S.; OLTJEN, J.W. STRATHE, A.B.; LANNA, D.P.D. A methodological approach to estimate the lactation curve and net energy and protein requirements of beef cows using nonlinear mixed-effects modeling. **Journal of Animal Science**, Champaign, v. 11, p. 3867-3878, Jun. 2012.

BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society**. London, v. 26, n. 1, p. 211-252, Apr. 1964.

DANTAS, O.D. **Avaliação de métodos para parametrização de modelos aplicados em sistemas agropecuários**. 2007. 85 p. Dissertação (Mestrado em Engenharia Elétrica e de Computação)-Universidade Federal de Goiás, Goiás.

FELÍCIO, P.E.; ALLEN, D.M.; CORTE, O.O. Influência na maturidade da carcaça sobre a qualidade de carne de novilhos Zebu. **Coletânea do Instituto de Tecnologia de Alimentos**, Campinas, v. 12, p. 137-150, 1981/1982.

HADDAD, C.M.; SANTOS, F.A.P. CASTRO, F.G.F. Avaliação da produção de novilho no Brasil. In: PEIXOTO, A.M. MOURA, J.C.; FARIA, V.P. (Ed.). **Confinamento de Bovinos**. Piracicaba: FEALQ, 2000. p. 21-44.

HANKINS, O.G.; HOWE, P.E. Estimation of the composition of beef carcass cuts. Tech. Bull. N. 26. USDA: Washington, DC, 1946.

LUCHIARI FILHO, A.; MOURA, A.C Influência do peso da carcaça e da espessura de gordura na maciez da carne bovina. **Pecuária de Corte**, Cuiabá, v. 8, n. 75, p. 56-58, 1998.

MARCONDES, M.I.; CHIZZOTTI, M.L.; VALADARES FILHO, S.C.; GIONBELLI, M.P.; PAULINO, P.V.R.; PAULINO, M.F. Energy requirements of zebu beef cattle. p. 81-106. 2010a. In: VALADARES FILHO, S.C.; MARCONDES, M.I.; CHIZZOTTI, M.L.; PAULINO, P.V.(Eds.) **Nutrient requirements of zebu cattle BR-CORTE**. 2.ed. DZO-UFV, Viçosa, MG, Brazil.

MARCONDES, M.I.; CHIZZOTTI, M.L.; VALADARES FILHO, S.C.; GIONBELLI, M.P.; PAULINO, P.V.R.; PAULINO, M. F. 2010b. Protein requirements of zebu beef cattle. p. 107-126. In: VALADARES FILHO, S.C.; MARCONDES, M.I.; CHIZZOTTI, M.L.; PAULINO, P.V. (Eds.) **Nutrient requirements of zebu cattle BR-CORTE**. 2.ed. DZO-UFV, Viçosa, MG, Brazil.

MCPHEE, M.J.; OLTJEN, J.W.; FADEL, J.G.; MAYER, D.G.; SAINZ, R.D. Parameter Estimation of Fat Deposition Models in Beef Steers. **DNA**, New Zelandia, v. 4, p. 115-120, Dec. 2007.

NRC. **Nutrient Requirements of Beef Cattle**. 7ed. National Academies Press, Washington, DC, 2000, 248 p.

OLTJEN, J.W.; PLEASANTS, A.B; SOBOLEVA, T.K. ODDY, V.H. Second-generation dynamic cattle growth and composition models. p. 197-209. 2000 In: MCNAMARA, J.P.; FRANCE, J.; BEEVER, D.E. **Modelling Nutrient Utilization in Farm Animals**. CAB International, London, LDN, United Kingdom.

OLTJEN, J.W.; BYWATER, A.C.; BALDWIN R.L.; GARRETT W.N. Development of a dynamic model of beef cattle growth and composition. **Journal of Animal Science**, Champaign, v. 62, p. 86-97, 1986.

PAULINO, P.V.R.; FONSECA, M.A.; HENRIQUES, L.T.; VALADARES FILHO, S.C.; DETMANN, E. Nutritional Requirements of nellore cows and calves. p. 167-185, 2010. In: VALADARES FILHO, S.C.; MARCONDES, M.I.; CHIZZOTTI, M.L.; PAULINO, P.V. (Eds.) **Nutrient requirements of zebu cattle BR-CORTE**. 2.ed. DZO-UFV, Viçosa, MG, Brazil.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014. Disponível em: <<http://www.R-project.org>>. Acesso em: 20 mar. 2014.

REID, J.T.; WELLINGTON, G.H.; DUNN H.O. Some relationships among the major chemical components of the bovine body and their application to nutritional investigations. **Journal of Dairy Science**, New York, v. 38, 1344-1359, 1955.

SAINZ, R.D. Qualidade das carcaças e da carne ovina e bovina. Curso sobre produção, qualidade e comercialização de carnes, v.1, Ribeirão Preto: USP, 1996.

SAINZ, R.D.; BARIONI, LG; PAULINO; P.V.; VALADARES FILHO, S.C.; OLTJEN, J.W. 2006. Growth patterns of Nellore vs British Beef Cattle Breeds assessed using a dynamic, mechanistic model of cattle growth and composition. p. 160-170. In: KEBREAB, E.; DIJKSTRA, J.; BANNINK, A.; GERRITS, W.J.J.; FRANCE, J. **Nutrient Digestion and utilization in Farm Animals: modeling approaches**. CAB International, London, LDN, United Kingdom.

SANTOS, F.A.P.; BALSALOBRE, M.A.A. Por que confinar?. In: PEIXOTO, A.M. MOURA, J.C.; FARIA, V.P. (Ed.). Confinamento de Bovinos. Piracicaba: FEALQ, 2000. p. 7-20.

VALENTE, E.E. L.; PAULINO, M.F.; DETMANN, E.; VALADARES FILHO, S.C.; CARDENAS, J.E.G.; DIAS, I.F.T. Requirement of energy and protein of beef cattle on tropical pasture. **Acta Scientiarum Animal Sciences**, Maringá, v. 35, p. 417-424, 2013.

4 PARAMETRIZAÇÃO DO MODELO DE CRESCIMENTO DE DAVIS PARA BOVINOS DE CORTE CRUZADOS

Resumo

O Modelo de Crescimento de Davis (MCD), um sistema de equações diferenciais que representa o crescimento de gado, foi parametrizado com dados de raças britânicas (*Bos taurus*) e utilizada com sucesso para a simulação e suporte a decisões nos Estados Unidos. No entanto, 30 anos após a parametrização original, é provável que o melhoramento genético e a adoção de diferentes raças tenham afetado os valores dos parâmetros do modelo. Portanto, os valores dos parâmetros podem precisar de ser re-estimados para animais Nelore (*Bos indicus*) e cruzados (*Bos taurus* × *Bos indicus*). Os objetivos deste estudo foram: calibrar o MCD usando a análise frequentista, o *bootstrap* não-paramétrico e simulações Monte Carlo; obter a distribuição assintótica dos parâmetros do MCD usando as duas últimas análises; comparar os resultados obtidos pelas três abordagens (frequentista, o *bootstrap* não-paramétrico e por Monte Carlo); comparar os resultados das melhores estimativas dos parâmetros propostas nesse presente estudo (animais de genótipos cruzados) com as estimativas originais do MCD obtidas pelo genótipo Britânico (*Bos taurus*); comparar os resultados obtidos a partir da abordagem da parametrização individuais com os intervalos de confiança obtidos da parametrização realizada na amostra total. Os valores dos parâmetros foram determinados usando o algoritmo de otimização *downhill simplex*, tanto nas análises frequentista como nas 300 reamostragem de *bootstrap* não-paramétrico. Para as simulações de Monte Carlo usou-se vetores de uma distribuição normal multivariada obtidos por meio do fator de Cholesky. Os principais critérios adotados para avaliar os ajustes do modelo foram o Erro Quadrático Médio de Predição (EQMP), o Coeficiente de Correlação de Concordância (CCC), o Critério de Informação Akaike (AIC) e o Critério de Informação Bayesiano (BIC). A técnica de *bootstrap* não-paramétrico com método de otimização *simplex downhill* estimou com sucesso as distribuições dos parâmetros (que tiveram distribuição probabilidade normal para maioria dos cenários). A taxa de deposição protéica apresentou um aumento mínimo de 156,17% e o máximo de 389,27% em comparação com os valores originais. O parâmetro relacionado a manutenção do animal teve um aumento mínimo de 125,81% e o máximo de 159,76% em comparação com os valores originais. As análises de *bootstrap* não-paramétrica produziram intervalos mais precisos para os Sistemas Intensivos e apresentou mais acurada e precisa para maioria dos cenários do experimento. Os limites inferiores dos intervalos de confiança com 95% para os parâmetros relacionados com a manutenção e as taxas de deposição protéica foram superiores às estimativas originais da MCD, evidenciando diferenças genéticas dos animais cruzados em relação aos parâmetros originais MCD. As tomadas de decisões baseadas nas análises individuais e na amostra total tiveram 50% de concordância. Concluiu-se que os genótipos cruzados tiveram maior gasto de energia de manutenção e foram mais rápidos de maturação em comparação tanto com os animais de genótipos Britânicos (*Bos Taurus*) e touros Nelores. As novas estimativas de valores de parâmetro, em vez dos originais, podem ser usadas para fins preditivos para Nelore e mestiços.

Palavras-chave: Otimização; Distribuição assintótica; *Bootstrap* não-paramétrico; Calibração; Não-linear

Abstract

The Davis Growth Model (DGM), a system of differential equations that represent cattle growth, has been parameterized with data from British (*Bos taurus*) breeds and successfully used for simulation and decision support in the United States. However, 30 years after the original parameterization, it is likely that genetic improvement and the adoption of different breeds affected model parameter values. Therefore, the parameter values may need to be re-estimated for Nellore (*Bos indicus*) and Nellore crossbred animals. The objectives of this study were: to calibrate the DGM using the frequentist analysis, nonparametric bootstrap and Monte Carlo methods; obtain the asymptotic distribution of DGM parameters using the last two analyzes; compare the results obtained by the three approaches (frequentist, nonparametric bootstrap and Monte Carlo); compare the results of the best estimates of the parameters proposed in the present study (animal crossed genotypes) with the original DGM estimates; compare the results obtained from the individual-based parameter fitting with the confidence intervals obtained from sample-based parameter fitting. The parameters values were determined using the downhill simplex optimization algorithm, in the frequentist analysis and in the 300 nonparametric bootstrap resampling. For the Monte Carlo simulations vectors were constructed from a multivariate normal distribution obtained by the Cholesky factor. Model accuracy was evaluated through the Mean Square Error of Prediction (MSEP), the concordance correlation coefficient (CCC), the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Nonparametric Bootstrap coupled with downhill simplex optimization successfully estimated parameters distribution, which did not deviated from normality for most scenarios. The protein deposition rate had a minimum increase of 156, 17% and the maximum of 389, 27% compared to the original values, and the maintenance requirement had a minimum increase of 125, 81% and the maximum of 159% compared to the original values. Analyses of nonparametric bootstrap produced more precise intervals for the Intensive Systems and was more accurate and precise for most of experimental scenarios. Lower limits of 95% confidence intervals for the parameters related to maintenance and protein accretion rates were higher than the original estimates of the DGM, evidencing genetic differences of the Zebu crossbred animals in relation to the original DGM parameters. Individual-based and the total sample analysis presented 50% agreement. It was concluded that cross-breed genotypes had higher maintenance energy expenditure and were faster maturing than original estimates for British breeds and Nellore and that the new parameter values estimates, instead the original ones, should be used for predictive purposes for Nellore and crossbred animals.

Keywords: Optimization; Asymptotic distribution; Nonparametric bootstrap; Calibration; Non-linear

4.1 Introdução

O comportamento das estruturas de correlações de parâmetros e suas estimativas para determinados cenários de crescimento de gado de corte podem ser interessante para adaptação de sistemas mais lucrativos e com menores impactos ambientais nos setores agropecuários. As matrizes de correlações dos parâmetros geradas pelos ajustes de sistemas dinâmicos com equações diferenciais ordinárias, por meio de técnica de *bootstrap* não-paramétrico e simulações Monte Carlo estabelecem a existência ou não da relação entre os parâmetros. A modelagem adequada para representar o crescimento do gado e dar sustentabilidade ao comportamento dos

animais é um importante requisito para obter melhores resultados e tomadas de decisões.

Funções de crescimento não-lineares (exemplo: Bhowmick & Bhattacharya, Gompertz, Logístico, Von Bertalanffy, Weibull and Richards) com parâmetros estimados por regressão descrevem o crescimento dos animais (Nesetrilová, 2005; Chizzotti et al., 2008; Forni et al., 2009; Freetly et al., 2011; Marinho et al., 2013; Bhowmick e Bhattacharya, 2014) mas tem capacidade limitada para estimar a composição corporal (isto é, gordura e proteína) e negligenciam as possíveis iterações no metabolismo e composição de ganho (NRC, 1996; Muniz et al., 2007).

Modelos matemáticos para crescimento de bovinos de corte como as propostas por Oltjen et al. (1986), France et al. (1987), Williams e Bennett (1995), Williams e Jenkins (2003), Hoch e Agabriel (2004) e Tedeschi et al. (2004) são potencialmente mais eficazes que as funções não-lineares mencionadas no parágrafo anterior, por expressar variáveis que incluem o metabolismo e a disponibilidade de nutrientes. O Modelo de Crescimento de Davis (MCD) descrito por Oltjen et al. (1986) e a atualização por Oltjen et al. (2000) foi escolhido para este estudo porque é um dos mais parcimoniosos e interpretável.

O MCD está atualmente sendo usado pelo programa predição fenotípica da Cooperative Research Centre (CRC) para predizer a gordura da 12 – 13^a costelas (*mm*) em bovinos de corte, Walmsley et al. 2014. A predição de gordura da 12 – 13^a costelas (*mm*) tem o potencial para auxiliar a indústria de carne bovina encontrar especificações rigorosas do mercado, tanto a nível nacional e internacional, que estão relacionados com o peso corporal (*kg*) e espessura de gordura subcutânea (*mm*). O gado frequentemente cae fora das especificações do mercado e portanto, os produtores são penalizados. Ao predizer a gordura da 12 – 13^a costela no momento do abate ajuda os produtores a cumprir rigorosas especificações do mercado e aumentar a rentabilidade (McPhee et al., 2007).

Embora existam vários estudos sobre exigências nutricionais de bovinos no Brasil (Sainz et al., 2006; Marcondes et al., 2010ab, b; Paulino et al., 2010; Valente et al., 2013), parametrizações do modelo proposto por Oltjen et al. (1986) para a genética zebuína (cruzados) exploradas em condições brasileiras nunca incluíram as estimativas da distribuição dos parâmetros por meio de inferências estatísticas como *bootstrap* não-paramétrico e simulações Monte Carlo.

Técnicas como de reamostragem *bootstrap* não-paramétrico vem sendo aplicado em diversas áreas para obter a distribuição assintótica dos parâmetros e inferências (intervalos de confiança, estimativas médias dos estimadores, moda, mediana e variância) principalmente quando não se conhece expressões fechadas da função de densidade dos parâmetros. Essa técnica corrobora na melhora da precisão e da acurácia suprindo situações como as de pequenas amostras, por exemplo, quando são comumente usadas nas áreas médicas e sistemas econômicos (Diebold et al., 1998; Matthews and Zhang, 2010; Belum, 2015). A enorme capacidade do cálculo dos computadores atuais facilita consideravelmente a aplicabilidade desta técnica, que é facilmente implementada. Essas técnicas também tem sido amplamente utilizadas em análise de incerteza (por exemplo, Efron, 1979; Srinivas e Srinivasan, 2005; Li et al., 2010).

Neste contexto, os objetivos deste estudo foram: calibrar o MCD usando a análise frequentista, o *bootstrap* não-paramétrico e simulações Monte Carlo, com dados nacionais; obter a distribuição assintótica dos parâmetros do MCD usando as duas últimas análises; comparar os resultados obtidos pelas três análises (frequentista, o *bootstrap* não-paramétrico e por simulações Monte Carlo); comparar os resultados das melhores estimativas dos parâmetros propostas nesse presente estudo (animais de genótipos cruzados) com as estimativas originais do MCD obtidas pela raça Britânica (*Bos taurus*); comparar os resultados obtidos a partir da abordagem dos estudos individuais (Capítulo 3) com os intervalos de confiança obtidos na amostra total (estudo desse presente Capítulo).

4.2 Material e métodos

Compôs-se o material e métodos em quatro subsecções: conjunto de dados, que descrevem o experimento; Modelo, com uma descrição das variáveis e parâmetros do MCD; Estatística Computacional, que descrevem os procedimentos dos algoritmos para resolver a EDO, os métodos usados para otimizar a função objetivo e obtenção das distribuições de probabilidade dos parâmetros, baseadas nas metodologias: frequentista, *bootstrap* não-paramétrico e simulação Monte Carlo; Avaliação de modelos, com as estatísticas usadas para comparar os ajustes.

4.2.1 Base de dados

O banco de dados do experimento nesse estudo foi conduzido em 2007 na Empresa Gado de Corte, descrito no Capítulo 3 da seção (3.2).

4.2.2 Modelo

O MCD, descrito no Capítulo 3 e seção (3.2), foi usado para obter as previsões baseadas nas amostras agregadas.

4.2.3 Estatística Computacional

Uma análise de variância multivariada preliminar, com as variáveis k_1 e α , calculadas simultaneamente pelas estimativas dos ajustes individuais do MCD foi realizada. O modelo estatístico utilizado na análise multivariada foi:

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\beta \times \gamma)_{jk} + \rho_{ijkl} + \epsilon_{ijkl}. \quad (4.1)$$

em que $i = 1$ e 2 referem aos genótipos, $1/2$ *Red Angus* $\times$ $1/4$ Caracu $\times$ $1/4$ Nelore (RCN) e *Red Angus* $\times$ Nelore (RN), respectivamente; $j = 1$ e 2 referem aos sistemas extensivo e intensivo, respectivamente; $k = 1$ e 2 referem aos gêneros fêmeas e machos não castrados, respectivamente; l são as repetições de cada tratamento, composto por um desbalanceamento, totalizando 27 unidades experimentais; $\mathbf{y}_{ijkl}$ é um vetor multivariado de observações para o i -ésimo genótipo, j -ésimo sistema, k -ésimo gênero e l -ésima repetição; $\boldsymbol{\mu}$ é um vetor de médias geral inerente a todas as observações; $\boldsymbol{\alpha}_i$ é um vetor de média de efeitos da i -ésimo genótipo; $\boldsymbol{\beta}_j$ é um vetor de média de efeitos do j -ésimo sistema; $\boldsymbol{\gamma}_k$ é um vetor de média de efeitos do k -ésimo gênero; $(\boldsymbol{\beta} \times \boldsymbol{\gamma})_{jk}$ é o vetor multivariado dos efeitos de interações entre o j -ésimo sistema e o k -ésimo gênero rebanho; $\boldsymbol{\rho}_{ijkl}$ é a covariável (vetor multivariado do efeito dos pesos iniciais associados às observações $\mathbf{y}_{ijkl}$); $\boldsymbol{\epsilon}_{ijk}$ é o vetor multivariado de erros associados às observações $\mathbf{y}_{ijkl}$.

Verificou-se que não havia diferença significativa dos vetores médios em relação aos genótipos. Procedeu-se as reamostragem de *bootstrap* não-paramétrico e simulações de MCMC usando o MCD considerando os ajustes populacionais baseados nas amostras dos grupos: fêmeas do sistema extensivo, fêmeas do sistema intensivo, machos do sistema extensivo e fêmeas do sistema intensivo. Detalhes de algumas estatísticas descritivas já foram apresentadas no Capítulo 3 nas Tabelas 3.1 e 3.2.

Análise Frequentista. O modelo foi implementado no *software* R (Development Core Team, versão 3.1, 2014). Para calcular a solução numérica de problemas de valores iniciais com sistemas de EDO de primeira ordem foi utilizado a função *ode()* do pacote deSolve (Soetaert et al., 2010). A solução foi realizada pelo método de integração *lsoda* com erro de tolerância absoluto e relativo de 10^{-6} . Para o cálculo da função objetivo que minimiza a soma de quadrados dos erros foi usada a seguinte equação

$$Res_{freq} = \sum_{i=1}^n \sum_{j=1}^p \left(\frac{\hat{y}_{ij} - y_{ij}}{\sigma_{ij}} \right)^2 \quad (4.2)$$

em que Res_{freq} representa a soma de quadrado residual de todas as observações da amostra; $\hat{y}_{ij}$ e y_{ij} são o valores preditos e observados, respectivamente, do i -ésimo indivíduo na j -ésima variável do modelo, $i = 1, 2, \dots, n$; $j = 1, \dots, p$; n representando o número de animais e p o número de variáveis respostas, compreendidas pela massa de proteína final e massa de gordura final; σ_{ij} é o desvio padrão do i -ésimo indivíduo na j -ésima variável do modelo. O desvio padrão é um fator de ponderação, que faz com que os termos se tornem adimensionais, portanto, ele é importante no caso de diferentes variáveis do modelo que tiverem diferentes unidades e magnitudes. Para obtenção dos vetores residuais e da soma de quadrados dos erros foi usada a função *modcost* do pacote FME, Soetaert e Petzoldt (2010). Os valores dos parâmetros foram determinados usando o algoritmo de otimização *downhill simplex* de Nelder e Mead (1965) disponível na função *modFit* do mesmo pacote. O método de Nelder-Mead otimiza problemas não-lineares e multidimensi-

onais, para isso requer somente a função objetivo e não precisa de derivadas para encontrar a solução. Esse método faz uso de um *simplex* inicial, isto é, um conjunto de vetores (pontos) num espaço M -dimensional, em que cada vértice representa uma solução possível (Press et al., 1990). O número de vértices *simplex* é igual ao número de parâmetros a ser ajustado, adicionado de um. Os coeficientes escala do método de Nelder-Mead para o fator de reflexão (β), o fator de contração (λ) e o fator de expansão (γ) receberam valores sugeridos por Press et al. (1990), 1,0, 0,5 e 2,0, respectivamente. O erro de tolerância foi 10^{-8} e o número máximo de interações foi tomado como 500.

Para o cálculo da matriz de covariância dos parâmetros foi considerada a aproximação numérica Jacobiana ($\mathbf{J}$), então, com base na teoria linear, a matriz de covariância foi estimada como (Press et al. 1992; Draper e Smith, 1998; Soetaert e Petzoldt, 2010):

$$\hat{V}(\hat{\theta}) = (\mathbf{J}^T \mathbf{J})^{-1} S^2 = \begin{bmatrix} \hat{V}(\hat{\theta}_1) & \hat{Cov}(\hat{\theta}_1, \hat{\theta}_2) & \cdots & \hat{Cov}(\hat{\theta}_1, \hat{\theta}_j) \\ \hat{Cov}(\hat{\theta}_1, \hat{\theta}_2) & \hat{V}(\hat{\theta}_2) & \cdots & \hat{Cov}(\hat{\theta}_2, \hat{\theta}_j) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{Cov}(\hat{\theta}_1, \hat{\theta}_j) & \hat{Cov}(\hat{\theta}_2, \hat{\theta}_j) & \cdots & \hat{V}(\hat{\theta}_j) \end{bmatrix} \quad (4.3)$$

com $j = 1, 2, 3, \dots, p$; p representando o número de parâmetros e $\hat{S}^2$ é a soma de quadrado residual do melhor ajuste. Com o melhor ajuste, $(\mathbf{J}^T \mathbf{J}) \approx 0,5 \mathbf{H}$, em que $\mathbf{H}$ é a matriz Hessiana.

Logo, o erro padrão da estimativa do parâmetro θ_j é dado por:

$$s(\hat{\theta}_j) = \sqrt{\hat{V}(\hat{\theta}_j)}. \quad (4.4)$$

Dessa forma, define-se o intervalo de confiança para θ_j como:

$$IC(\theta_j) : \hat{\theta}_j \pm t_{(G.L.erro, \alpha/2)} \sqrt{\hat{V}(\hat{\theta}_j)}. \quad (4.5)$$

Bootstrap não-paramétrico. Foram realizadas para cada cenário, 300 reamostragem de *bootstrap* não-paramétrico e calculado o intervalo de confiança *Biased Corrected Percentile Bootstrap* (BCPB). A distribuição de probabilidades dos parâmetros, k_1 e α , foi representada graficamente.

Todos os cálculos realizados para as reamostragens de *bootstrap* não-paramétrico foram semelhantes aos realizados pela análise frequentista, porém, sendo realizados para os conjuntos de reamostragens realizadas com reposição por meio da amostra original. Ou seja, para cálculo da solução numérica de problemas de valores iniciais com sistemas de EDO de primeira ordem foi usada a função *ode()* do pacote deSolve (Soetaert et al., 2010). A solução foi realizada pelo método de integração *lsoda* com erro de tolerância absoluto e relativo de 10^{-6} . Ao final

das 300 reamostragens, foi calculado o intervalo de confiança BCPB, obtido conforme descrito a seguir.

Intervalo de Confiança BCPB. Para a construção do intervalo de confiança BCPB foi ordenado de forma crescente as B estimativas para cada reamostragem, representada por $\hat{\theta}_b^*$ e calculado as proporções de réplicas *bootstrap* menores que o valor da estatística original, $\hat{\theta}$, ou seja,

$$p_0 = P[\hat{\theta}_b^* \leq \hat{\theta}] \quad (4.6)$$

com $b = 1, 2, 3, \dots, B$, sendo B o número de reamostragens realizadas. O asterístico é usado para diferenciar as estimativas das reamostras e as estimativas da amostra original. A partir do valor encontrado p_0 foi obtido o parâmetro de correção do viés que é definido por

$$Z_0 = \Phi^{-1}(p_0), \quad (4.7)$$

sendo $\Phi(\bullet)$ a função de distribuição acumulada normal padrão. Assim, as correções propostas neste método para o percentil inferior e superior foram

$$P_I = \Phi[2Z_0 - Z_{\alpha/2}] \text{ e } P_S = \Phi[2Z_0 + Z_{\alpha/2}] \quad (4.8)$$

e o intervalo de confiança *bootstrap* BCPB é dado por

$$IC_{BCPB} = [\theta_{(P_I)}^*; \theta_{(P_S)}^*], \quad (4.9)$$

em que $\theta_{(P_I)}^*$ e $\theta_{(P_S)}^*$ representam respectivamente, o P_I -ésimo percentil e o P_S percentil das reamostras de *bootstrap* não-paramétrico.

Para o cálculo da função objetivo que minimiza a soma de quadrados dos erros foi usada a seguinte equação

$$Res_{boot} = \sum_{i=1}^n \sum_{j=1}^p \left(\frac{\hat{y}_{ij} - y_{ij}}{\sigma_{ij}} \right)^2 \quad (4.10)$$

em que Res_{boot} representa a soma de quadrado residual de todas as observações da reamostra; $\hat{y}_{ij}$ e y_{ij} são os valores preditos e observados, respectivamente, do i -ésimo indivíduo na j -ésima variável do modelo, $i = 1, 2, \dots, n$; $j = 1, \dots, p$; n representando o número de animais e p o número de variáveis respostas, compreendidas pela massa de proteína final e massa de gordura final; σ_{ij} é o desvio padrão do i -ésimo indivíduo na j -ésima variável do modelo.

Os valores dos parâmetros foram determinados usando o algoritmo de otimização *downhill simplex* de Nelder e Mead (1965) disponível na função *optim* do pacote *stats* (R Core Team, 2014). Os coeficientes escala do método de Nelder-Mead para o fator de reflexão (β), o fator de contração (λ) e o fator de expansão (γ) receberam valores sugeridos por Press et al. (1990),

1,0, 0,5 e 2,0, respectivamente. O erro de tolerância foi 10^{-8} e o número máximo de interações foi tomado como 500.

Simulações Monte Carlo. Amostras aleatórias normais multivariadas de tamanho n foram simuladas no espaço p -dimensional, dadas por $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$. O vetor aleatório $\mathbf{X}_j = [X_{j1}, X_{j2}, \dots, X_{jp}]^t$ de dimensão $p \times 1$ tem densidade normal multivariada dada por

$$f(\mathbf{x}) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\} \quad (4.11)$$

em que Σ é a matriz de covariâncias e $\boldsymbol{\mu}$ é o vetor de médias $p \times 1$ populacionais. Para simular um vetor aleatório $\mathbf{X}_j$ da distribuição apresentada em (4.11) foi obtida o fator de Cholesky (Γ) da matriz Σ de dimensão $p \times p$, tal que $\Sigma = \Gamma \Gamma^t$. Posteriormente, foi simulado um vetor $\mathbf{Z}_j$ composto de elementos independentes de uma distribuição normal padrão univariada, sendo que $\mathbf{Z}_j$ tem distribuição $N_p(\mathbf{0}, \mathbf{I})$, em que $\mathbf{0}$ é o vetor nulo e $\mathbf{I}$ é a matriz identidade $p \times p$. Utilizando uma transformação linear do vetor dada por $\mathbf{X}_j = \Gamma \mathbf{Z}_j + \boldsymbol{\mu}$, obteve-se o vetor desejado com distribuição dada por (4.11). Repetiu-se esse processo $n = 5.000$ vezes, até formar a amostra aleatória. Para cada cenário (fêmeas do sistema intensivo, fêmeas do sistema extensivo, machos do sistema extensivo e machos do sistema intensivo) as amostras foram compostas pelos vetores de médias e matrizes de covariâncias obtidos pelos resultados de *bootstrap* não-paramétrico de cada cenário. Intervalos percentílicos foram obtidos para representar as informações da populações de cada cenário.

4.2.4 Avaliação de modelos

Todas as estatísticas, descritas no Capítulo 2 e seção (2.2.8), foram usadas para avaliar o MCD, baseadas nos ajustes sobre a abordagem frequentista, *bootstrap* não-paramétrico e simulações Monte Carlo.

4.3 Resultados e Discussões

Os resultados das novas parametrizações juntamente com os intervalos de confiança, usando análise frequentista e reamostragem por *bootstrap* não-paramétrico são apresentadas na Tabela 4.1. Intervalos mais precisos foram verificados para todos os cenários sob os ajustes com o método de *bootstrap* não-paramétrico. O teste de Kolmogorov-Smirnov (KS) mostrou distribuição de normalidade para todos os parâmetros (valor $p > 0,05$), exceto para k_1 para fêmeas no sistema extensivo. Isso mostra que as distribuições de probabilidades não são as mesmas evidenciando diferenças entre os grupos, como fêmeas no sistema extensivo e machos para ambos os sistema oriundos do período antes o desmame quanto ao parâmetro de acréscimo de DNA (k_1).

As matrizes de dispersões conjuntas de estimação *bootstrap* não-paramétricos, a correlação de *Pearson* e os histogramas de frequências para k_1 e α são mostrados na Figura 4.1. Verificou-se também que não houve diferença entre os sistemas extensivo e intensivo dentro do gênero fêmeas, considerando os parâmetros k_1 e α simultaneamente, pelas duas metodologias frequentista e por reamostragem *bootstrap*.

Tabela 4.1 – Estimativas e Intervalos de confiança frequentistas e *bootstrap* não-paramétrico para o MCD

Modelo	constante fixada	Sistema	Parâmetros	Frequentista		<i>Bootstrap</i>		D.P	valor <i>p</i> KS	Sobreposição
				Média Freq.	I.C.	Média <i>Bootstrap</i>	I.C. BCPC			
Fêmeas	$(k_2 = 0, 0461)^1$ $(k_3 = 0, 1430)^1$	Extensivo	k_1	0,0160	(0,0070; 0,0246)	0,0167	(0,0136; 0,0187)	0,0018	0,0460	28,91%
			α	0,1100	(0,0983; 0,1218)	0,1082	(0,0993; 0,1185)	0,0065	0,0894	81,70%
	$(DNA_{max} = 308)^2$	Intensivo	k_1	0,0119	(0,0057; 0,01799)	0,0119	(0,0079; 0,0160)	0,0020	0,7173	66,13%
			α	0,1216	(0,1127; 0,1304)	0,1214	(0,1109; 0,1302)	0,0022	0,9961	43,36%
Machos	$(k_2 = 0, 0461)^1$ $(k_3 = 0, 1430)^1$	Extensivo	k_1	0,0067	(0,0044; 0,0095)	0,0067	(0,0035; 0,0095)	0,0017	0,1882	84,21%
			α	0,1371	(0,1236; 0,1489)	0,1374	(0,1311; 0,1389)	0,0024	0,1595	30,83%
	$(DNA_{max} = 462)^2$	Intensivo	k_1	0,0099	(0,0075; 0,0114)	0,0096	(0,0086; 0,0108)	0,0005	0,5842	56,26%
			α	0,1203	(0,1136; 0,12652)	0,1204	(0,1131; 0,1270)	0,0031	0,2023	82,73%

¹Valores fixados baseados em Oltjen et al. (1986); ² DNA_{max} foi baseado na proporção direta e linear do peso maduro dos referidos genótipos e gêneros desses animais.

Na metodologia *bootstrap* não-paramétrico, correlações positiva entre esses parâmetros k_1 e α foram verificadas para os grupos: fêmeas do sistema intensivo, fêmeas do sistema extensivo e machos do sistema intensivo, esse último caso apresentou uma fraca correlação positiva. Uma correlação negativa foi encontrada para machos não castrados do sistema extensivo, indicando que os animais crescem mais, ou seja, depositam proteína mais rápido e são mais eficientes no uso da energia, Figura 4.1.

Os resultados das simulações Monte Carlo apresentaram resultados médios e desvio padrão similares aos da distribuição teórica de *bootstrap* não-paramétrico. Os intervalos de confiança da simulação Monte Carlo apresentaram pequenas variações em relação aos resultados encontrados pelo *bootstrap* na Tabela 4.2, com sobreposição de todos os parâmetros e cenários analisados. As matrizes de correlação dos parâmetros foram praticamente similares entre as duas metodologias (*bootstrap* não-paramétrico e simulações Monte Carlo). Os gráficos das dispersões dos parâmetros para cada cenário com as simulações de 5.000 vetores de k_1 e α , baseadas nas correlações de *bootstrap* não-paramétrico estão apresentadas na Figura 4.2. As simulações de Monte Carlo suavizaram as densidades das distribuições dos parâmetros e a medida que o número de simulações aumentam espera-se aproximar da densidade teórica da distribuição normal.

Houve sobreposição dos intervalos de confiança obtidos pela análise frequentista, os intervalos de confiança obtidos pela análise de *bootstrap* não-paramétrico e os intervalos de confiança obtidos pela análise por simulação Monte Carlo, em todos os casos. Os intervalos de confiança obtidos pelas análises Monte Carlo e *bootstrap* não-paramétrico mostraram que as taxas de energia de manutença (α) dos machos extensivos foram diferentes em relação aos demais cenários desse estudo. Não houve diferença quanto a manutença entre machos e fêmeas do sistema intensivo.

Figura 4.1 – Dispersão conjunta, correlação de *Pearson* e histograma de frequência para k_1 e α para o MCD por meio da análise de *bootstrap* não-paramétrico

Figura 4.2 – Dispersão conjunta, correlação de *Pearson* e histograma de frequência para k_1 e α para o MCD por meio de simulações Monte Carlo

Tabela 4.2 – Medidas descritivas, teste Kolmogorov-Smirnov (valor p) e intervalos de confiança para as médias dos parâmetros com 95% de confiança baseadas nas Simulações Monte Carlo

Gênero	Sist. pré-desmame	constantes fixas	Par.	Simulações MC ¹		valor <i>p</i> KS	IC		Sobreposição (<i>bootstrap</i> e MC)
				Média	D.P.		LI	LS	
Fêmeas	Extensivo	(<i>k</i> ₂ = 0, 0461) ²	<i>k</i> ₁	0,0167	0,0018	0,7301	0,0131	0,0201	72,85%
		(<i>k</i> ₃ = 0, 1430) ²	<i>α</i>	0,1082	0,0065	0,5390	0,0952	0,1209	74,71%
		(<i>DNA</i> _{<i>max</i>} = 308) ³							
	Intensivo	(<i>k</i> ₂ = 0, 0461) ²	<i>k</i> ₁	0,0119	0,0020	0,6513	0,0081	0,0158	95,06%
(<i>k</i> ₃ = 0, 1430) ²		<i>α</i>	0,1214	0,0022	0,8654	0,1170	0,1256	97,67%	
(<i>DNA</i> _{<i>max</i>} = 308) ³									
Machos	Extensivo	(<i>k</i> ₂ = 0, 0461) ²	<i>k</i> ₁	0,0067	0,0017	0,4582	0,0034	0,0099	92,30%
		(<i>k</i> ₃ = 0, 1430) ²	<i>α</i>	0,1374	0,0024	0,9741	0,1325	0,1420	58,71%
		(<i>DNA</i> _{<i>max</i>} = 462) ³							
	Intensivo	(<i>k</i> ₂ = 0, 0461) ²	<i>k</i> ₁	0,0096	0,0005	0,5626	0,0086	0,0105	86,36%
(<i>k</i> ₃ = 0, 1430) ²		<i>α</i>	0,1204	0,0031	0,3324	0,1138	0,1263	92,00%	
(<i>DNA</i> _{<i>max</i>} = 462) ³									

¹MC: Monte Carlo; ²Valores fixados baseados em Oltjen et al. (1986); ³ DNA_{max} foi baseado proporção direta e linear do peso maduro das referidos genótipos e gêneros desses animais.

Comparações foram realizadas para verificar o desempenho das estimativas frequentista, *bootstrap* e as simulação Monte Carlo, Tabela 4.3. As avaliações do *bootstrap* e das estimativas médias realizadas pelas simulações de Monte Carlo tiveram resultados do desempenho igual para o MCD. Para o cenário fêmeas nos sistemas extensivo não houve uma melhor metodologia que se ajustasse as estimativas de proteína e gordura simultaneamente. Por exemplo, observou-se que a análise frequentista o MCD ajustou melhor as mensurações de proteína (menores valores de EQMP, MB, AIC, BIC e maiores valores para r^2 , MEF, CD, C_b e CCC) e as análises de *bootstrap* e das simulações Monte Carlo foram melhores para ajustar a gordura (menores valores de EQMP, MB, AIC, BIC e maiores valores para r^2 , MEF, C_b e CCC).

Tabela 4.3 – Avaliação do MCD com as estimativas obtidas pelas análises frequentistas, *bootstrap* e simulação Monte Carlo para quatro cenários de estudos (fêmeas e machos nos sistemas extensivos e intensivos)

Cenários	mensurações	Frequentista		<i>Bootstrap</i>		Monte Carlo	
		Proteína	Gordura	Proteína	Gordura	Proteína	Gordura
Fêmeas Sistema Extensivo	Média	59,06568	72,4256	59,2117	73,0537	59,2117	73,0537
	SD	3,795524	9,909386	3,8113	10,1575	3,8113	10,1575
	Mediana	59,68831	73,16988	59,8289	73,7612	59,8289	73,7612
	r^2	0,9314	0,7131	0,8916	0,5759	0,8916	0,5759
	EQMP	1,77293	25,8227	2,1369	24,892	2,1369	24,892
	MB	-1,0156	1,9743	-1,1617	1,3462	-1,1617	1,3462
	MEF	0,8272	0,3602	0,7918	0,3833	0,7918	0,3833
	CD	0,8670	0,5205	0,8381	0,5096	0,8381	0,5096
	C_b	0,9642	0,9325	0,9539	0,9384	0,9539	0,9384
	CCC	0,9386	0,8262	0,9270	0,8352	0,9270	0,8352
	valor p^1	0,8370	0,2904	0,8283	0,2753	0,8283	0,27535
	valor p^2	0,7732	0,1555	0,7576	0,2889	0,7576	0,2889
	valor p^3	0,5679	0,6191	0,5150	0,6286	0,5150	0,6286
	AIC	15,9431	27,1482	16,1524	27,0901	16,1524	27,0901
	BIC	14,1020	25,3071	14,3113	25,2490	14,3113	25,2490
Fêmeas Sistema Intensivo	Média	56,0478	67,8593	56,0570	67,9391	56,0570	67,9391
	SD	2,3111	7,5987	2,3119	7,6045	2,3119	7,6045
	Mediana	56,3338	67,9320	56,3446	68,0205	56,3446	68,0205
	r^2	0,8870	0,7720	0,8860	0,7710	0,8860	0,7710
	EQMP	7,5780	17,1615	7,6008	17,0037	7,6008	17,0037
	MB	-1,1701	1,5185	-1,1792	1,4386	-1,1792	1,4386
	MEF	0,6140	0,1580	0,6130	0,1650	0,6130	0,1650
	CD	3,2130	0,3800	3,2000	0,3810	3,2000	0,3810
	C_b	0,7540	0,8770	0,7100	0,7540	0,7100	0,7540
	CCC	0,7100	0,7710	13,2310	0,7720	13,2310	0,7720
	valor p^1	0,0085	0,0045	0,0086	0,0046	0,0086	0,0046
	valor p^2	0,0094	0,0056	0,0095	0,0056	0,0095	0,0056
	valor p^3	0,0249	0,0190	0,0249	0,0199	0,0249	0,0199
	AIC	38,7368	45,3435	38,7718	45,3893	38,7718	45,3893
	BIC	39,3285	45,9351	39,3634	45,9810	39,3634	45,9810
	AIC_c	13,196	19,8030	13,2310	19,8480	13,2310	19,8480
Machos Sistema Extensivo	Média	76,0336	72,9817	75,4652	72,8551	75,4652	72,8551
	SD	5,2796	17,8191	5,2754	17,7340	5,2754	17,7340
	Mediana	75,9689	66,3573	75,5530	66,2709	75,5530	66,2709
	r^2	0,7140	0,9050	0,7290	0,9040	0,7290	0,9040
	EQMP	20,9197	41,6166	19,6221	40,9915	19,6221	40,9915
	MB	-0,9136	2,1183	-0,3452	2,2448	-0,3452	2,2448
	MEF	0,6470	0,6900	0,6690	0,6950	0,6690	0,6950
	CD	2,5600	0,5200	2,6480	0,5260	2,6480	0,5260
	C_b	0,8840	0,9430	0,8900	0,9430	0,8900	0,9430
	CCC	0,7470	0,8960	0,7600	0,8970	0,7600	0,8970
	valor p^1	0,4958	0,0853	0,4784	0,0878	0,4784	0,0878
	valor p^2	0,5074	0,0985	0,4822	0,1030	0,4822	0,1030
	valor p^3	0,8191	0,2810	0,8298	0,2844	0,8298	0,2844
	AIC	34,3288	32,9886	34,0756	32,9613	34,0756	32,9613
	BIC	33,1571	31,81693	32,9039	31,8005	32,9039	31,8005
	AIC_c	24,139	22,7990	23,8860	22,7720	23,8860	22,7720
Machos Sistema Intensivo	Média	79,0522	76,0969	77,2364	77,6431	77,2364	77,6431
	SD	3,7913	8,9600	3,6848	8,8158	3,6848	8,8158
	Mediana	77,2088	78,2749	75,4895	79,5552	75,4895	79,5552
	r^2	0,7450	0,4410	0,7900	0,4690	0,7900	0,4690
	EQMP	15,5610	47,8119	6,3603	38,2956	6,3603	38,2956
	MB	-3,2411	2,8031	-1,4253	1,2569	-1,4254	1,2569
	MEF	0,2050	-0,4150	0,6750	-0,1340	0,6750	-0,1340
	CD	0,8410	0,4260	1,3890	0,4780	1,3890	0,4780
	C_b	0,7590	0,8760	0,9190	0,9270	0,9190	0,9270
	CCC	0,6550	0,5810	0,8170	0,6340	0,8170	0,6340
	valor p^1	0,6578	0,0210	0,5182	0,0275	0,5182	0,0275
	valor p^2	0,7807	0,0268	0,5691	0,0304	0,5691	0,0304
	valor p^3	0,0332	0,0643	0,2907	0,1094	0,2907	0,1094
	AIC	46,0200	57,9913	44,2809	57,5284	44,2809	57,5284
	BIC	46,6117	58,5830	44,8725	58,1201	44,8725	58,1201
	AIC_c	20,4790	32,4500	18,7400	31,9870	18,7400	31,9870

valor p^1 : ($H_0: \beta_0 = 0$); valor p^2 : ($H_0: \beta_1 = 1$) e valor p^3 : ($H_0: \beta_0 = 0$ e $\beta_1 = 1$).

O desempenho das estimativas dos resultados das análises frequentistas foram me-

lhores para o cenário das fêmeas do sistema intensivo. Para os cenários machos de ambos os sistemas, as análises *bootstrap* e simulações Monte Carlo tiveram desempenho melhor para representar tanto a proteína quanto a gordura (menores valores de EQMP, MB, AIC, BIC e maiores valores para r^2 , MEF, CD, C_b e CCC). O r^2 , apesar de não ser um critério adequado para verificar a qualidade de ajustamento, conforme discutido no capítulo 2, apresentou altos valores para o ajuste da proteína ($r^2 > 0,7131$).

O modelo reparametrizado teve as predições melhoradas usando as quatro metodologias do estudo. Um esboço comparativo das estimativas obtidas pelas análises de *bootstrap*, frequentistas as parametrizações originais propostas por Oltjen et al. (1986) mostram que as calibrações realizadas nesse estudo são mais adequadas para representar o rebanho do genótipo cruzado (Figura 4.3).

Uma análise feita por McPhee et al. (2007) indicou que o modelo havia sobrestimado o ganho de gordura para todos os depósitos de gordura (subcutânea, intramuscular, visceral e intermuscular). As avaliações feitas nesse estudo com animais cruzados com as estimativas médias obtidas tanto pela metodologia *bootstrap* como para a metodologia frequentista, mostra que o MCD sobrestimou a proteína e subestimou o depósito de gordura dos animais (Tabela 4.4). Os vieses calculados em relação proteína e gorduras médias usando as novas parametrizações propostas nesse estudo em relação a proteína e gordura média observadas são menores do que os vieses calculados de proteína e gordura usando a calibração original do modelo Oltjen et al. (1986), respectivamente. Portanto, este modelo ajustado foi aceito provisoriamente para a simulação do crescimento de genótipos de animais Nelore (*Bos indicus*) reparametrizações realizadas por Sainz et al. (2006) e cruzados (do presente estudo), enquanto se aguardam novas avaliações com um conjunto de dados independente.

Tabela 4.4 – Comparações das estimativas médias de proteína e gordura entre as novas calibrações propostas nesse estudo e as calibrações usando os valores de Oltjen et al. (1986)

Grupos	Variáveis	Observado	Novas calibrações <i>bootstrap</i>		Novas calibrações Frequentista		Calibração original	
			Predições	Viés	Predições	Viés	Predições	Viés
Fêmeas	Proteína	58,0500	59,2117	-1,1617	59,0656	-1,0156	53,8236	-4,2264
Sistema Extensivo	Gordura	74,4000	73,0537	1,3462	72,4256	1,9743	87,3991	12,9991
Fêmeas	Proteína	54,8777	56,0579	-1,1792	56,0478	-1,1701	51,9247	2,9530
Sistema Intensivo	Gordura	69,3700	67,9391	1,4386	67,8593	1,5185	82,7235	-13,3535
Machos	Proteína	75,1200	75,4652	-0,3452	76,0336	-0,9136	53,8236	5,5314
Sistema Extensivo	Gordura	75,1000	72,8551	2,2448	72,9817	2,1183	87,3991	-39,3059
Machos	Proteína	75,8111	77,2364	-1,4253	77,2364	-3,2411	67,8307	7,9804
Sistema Intensivo	Gordura	78,9000	77,6431	1,2568	77,6431	2,8031	105,9546	-27,0546

Comparações usando as parametrizações do *bootstrap* não-paramétrico em relação aos valores de k_1 e α foram maiores do que os parâmetros originais do MCD proposto por Oltjen et al. (1986) e também em relação a reparametrização realizada para Sainz et al. (2006). No estudo Oltjen et al. (1986), k_1 teve um aumento mínimo de 156, 17% e máximo de 389, 27% em

Figura 4.3 – Valores observados e estimados para as proteínas e as gorduras - modelo DGM ajustando com os parâmetros estimados da proposta desse estudo (à direita) e com os parâmetros originais de referência de touros *Bos Taurus* (à esquerda)

comparação com os valores originais. O parâmetro α teve um aumento mínimo de 125,81% e o máximo de 159,76% em relação aos valores iniciais (Tabela 4.5).

Nos estudos anteriores envolvendo reparametrização do modelo com touros do genótipo Nelore por Sainz et al. (2006), estimavas foram realizadas com as taxas de k_1 , k_3 e

Tabela 4.5 – Porcentagem de mudança dos atuais valores dos parâmetros calibrados para genótipos cruzados (por *Bootstrap*) em relação aos valores dos parâmetros originais do MCD com genótipo Britânico *Bos taurus* realizado por Oltjen et al. (1986) e com a parametrização de Nelore *Bos Indicus* realizado por Sainz et al. (2006)

<i>Bootstrap</i> Parâmetros ajustado	Genótipo Cruzado				Britânico <i>Bos taurus</i> Novilhos	Nelores <i>Bos indicus</i> Touros
	Fêmeas SE ¹	Fêmeas SI ²	Machos SE	Machos SI		
k_1	0,0167	0,0119	0,0067	0,0096	0,00429 [‡]	0,00304 [§]
α	0,1082	0,1214	0,1374	0,1204	0,0860 [‡]	0,0768 [§]
% de mudança em relação calibração original referência Oltjen et al. (1986)						
k_1	k_1 (389,27%)	k_1 (277,38%)	k_1 (156,17%)	k_1 (223,77%)	***	***
α	α (125,81%)	α (141,16%)	α (159,76%)	α (140,00%)	***	***
% de mudança em relação calibração original referência Sainz et al. (2006)						
k_1	k_1 (549,34%)	k_1 (391,44%)	k_1 (220,39%)	k_1 (315,78%)	***	***
α	α (140,88%)	α (158,07%)	α (178,90%)	α (156,77%)	***	***

¹Sistema Extensivo; ²Sistema Intensivo; [‡] Parâmetros estimados para o genótipo Britânico *Bos taurus* por Oltjen et al. (1986); [§] Parâmetros estimados para o genótipo Nelore *Bos Indicus* por Sainz et al. (2006).

α , mantendo k_2 (0,0479) e DNA_{max} (462) fixas. Neste estudo, k_1 teve um aumento mínimo de 220,39% e o máximo de 549,34%; o α teve um aumento mínimo de 140,88% e o máximo de 178,90% em relação ao genótipo *Bos indicus* (Tabela 4.5).

Os intervalos de confiança para os parâmetros dos animais mestiços neste estudo não inclui os parâmetros do genótipo Nelore encontrado por Sainz et al. (2006). Nesse estudo anterior, os autores indicaram que animais do genótipo Nelore é mais lento de maturação, têm menores taxas de retorno da proteína e têm menores gastos de energia endógena (Sainz et al., 2006). Os resultados indicam o oposto para os animais mestiços do nosso conjunto de dados, ou seja, gastos superiores com energia endógenas e animais mais rápidos na maturação, em comparação tanto com Oltjen et al. (1986) e Sainz et al. (2006).

O gráfico dos resíduos estudentizados em função da ordem das observações demonstra a qualidade do ajuste proporcionado para cada cenário. Caso o resíduo apresente-se distribuído aleatoriamente em relação a ordem das observações é indicativo de que o ajuste foi adequado. Conforme a inspeção visual dos gráficos de dispersão dos resíduos (Figura 4.4), observa-se que os modelos se adequaram para cada grupo. Os testes foram realizados para comprovar as pressuposições de normalidade (teste de Shapiro-Wilk), homogeneidade (teste de Breusch-Pagan) e independência (teste de Durbin-Watson) dos resíduos, Tabelas 4.6 e 4.7.

Tabela 4.6 – Análise residual para cada compartimento das variáveis de estado do MCD expresso pelo valor p dos testes Breusch-Pagan, Durbin-Watson e Shapiro-Wilk, baseadas na estimativas de *bootstrap* não-paramétrico

Gênero	Sistema	Variável	Homogeneidade Breusch-Pagan	Independência Durbin-Watson	Normalidade Shapiro-Wilk
Fêmeas	Extensivo	Proteína	0,5624	0,6560	0,9495
	Extensivo	Gordura	0,7367	0,2440	0,5546
	Intensivo	Proteína	0,1942	0,3420	0,9281
	Intensivo	Gordura	0,5417	0,8100	0,8880
Machos	Extensivo	Proteína	0,1516	0,7560	0,2531
	Extensivo	Gordura	0,2214	0,5300	0,4915
	Intensivo	Proteína	0,3869	0,5560	0,8982
	Intensivo	Gordura	0,9018	0,2040	0,5014

Tabela 4.7 – Análise residual do MCD expresso pelo valor p dos testes Breusch-Pagan, Durbin-Watson e Shapiro-Wilk, baseadas na estimativas de *bootstrap* não-paramétrico

Gênero	Sistema	Homogeneidade Breusch-Pagan	Independência Durbin-Watson	Normalidade Shapiro-Wilk
Fêmeas	Extensivo	0,0900	0,5520	0,6607
	Intensivo	0,8600	0,2760	0,9927
Machos	Extensivo	0,5213	0,5320	0,7299
	Intensivo	0,3869	0,6560	0,2010

Figura 4.4 – Dispersão dos resíduos estudatizados em relação a ordem das observações para os cenários do experimento parametrização bootstrap não-paramétrico do MCD

4.3.1 Comparações da abordagem baseadas nos estudos individuais e na amostra total

O resumo de todas as metodologias comparativas envolvendo intervalos de confiança e testes de hipóteses, com nível de significância de 5%, baseadas nos ajustes de soma de quadrados individuais (Capítulo 3) e da amostra (Capítulo 4) foram apresentadas na Tabela 4.8. Os resultados obtidos utilizando as três metodologias (frequentista, *bootstrap* não-paramétrico e simulações Monte Carlo) por meio da amostra total (estudo desse presente Capítulo) tiveram resultados semelhantes quanto as tomadas de decisões, baseadas nas hipóteses apresentadas na Tabela 4.8. As decisões baseadas nas análises individuais e na amostra total baseados nos *bootstrap* não-paramétrico tiveram 50% de concordância, já as análises individuais e na amostra total baseadas nos intervalos de confiança frequentista tiveram 28% de acertos. Ao realizar um teste de hipótese ou um intervalo de confiança, o interesse maior do pesquisador está na tomada de decisão quanto a não rejeição ou rejeição da hipótese referente ao parâmetro populacional (Steel et al., 1997). Desse modo, ao testar uma hipótese de nulidade, ele deve levar em consideração o risco global de tomar decisões erradas e está sujeito a incorrer nos possíveis erros (Mood et al., 1974; Steel et al., 1997; Baron, 2006; Hogg e Tanis, 2010):

1. Erro tipo *I*: erro cometido ao se rejeitar a hipótese nula verdadeira, quando deveria ser aceita. Esse erro é controlado diretamente pelo pesquisador e a probabilidade de se cometê-lo é dada pelo nível de significância α , ou seja, $P[\text{Erro Tipo } I] = P[\text{rejeitar } H_0 | H_0 \text{ verdadeira}] = \alpha$;
2. Erro tipo *II*: erro cometido ao não se rejeitar a hipótese nula falsa. Esse erro não é controlado diretamente pelo pesquisador e a probabilidade de se cometer esse erro, denotada por β , é dada por $P[\text{Erro Tipo } II] = P[\text{não rejeitar } H_0 | H_0 \text{ falsa}] = \beta$.

Tabela 4.8 – Comparações das hipóteses baseados nos ajustes individuais e nos ajustes baseados na amostra total baseado no MCD, Oltjen et al. (1986), ao nível de 5% de significância

parâmetro	Hipóteses	Decisão baseada na hipótese estatística			
		Abordagens metodológicas			
		indivíduo	amostra (lotes de animais)		
		ANOVA	frequentista	<i>bootstrap</i>	Monte Carlo
k_1	$H_0: \mu_{FSI}^1 + \mu_{FSE}^2 = \mu_{MSI}^3 + \mu_{MSE}^4$	Rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
	$H_0: \mu_{FSI} = \mu_{FSE}$	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
	$H_0: \mu_{MSI} = \mu_{MSE}$	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
	$H_0: \mu_{FSE} = \mu_{MSI}$	Rejeita H_0	Não rejeita H_0	Rejeita H_0	Rejeita H_0
	$H_0: \mu_{FSI} = \mu_{MSE}$	Rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
	$H_0: \mu_{FSI} = \mu_{MSI}$	Rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
	$H_0: \mu_{FSE} = \mu_{MSE}$	Rejeita H_0	Não rejeita H_0	Rejeita H_0	Rejeita H_0
	$H_0: \mu_{FSI} + \mu_{FSE} = \mu_{MSI} + \mu_{MSE}$	Rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
α	$H_0: \mu_{FSI} = \mu_{FSE}$	Rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
	$H_0: \mu_{MSI} = \mu_{MSE}$	Rejeita H_0	Não rejeita H_0	Rejeita H_0	Rejeita H_0
	$H_0: \mu_{FSE} = \mu_{MSI}$	Rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
	$H_0: \mu_{FSI} = \mu_{MSE}$	Não Rejeita H_0	Não rejeita H_0	Rejeita H_0	Rejeita H_0
	$H_0: \mu_{FSI} = \mu_{MSI}$	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
	$H_0: \mu_{FSE} = \mu_{MSE}$	Rejeita H_0	Não rejeita H_0	Rejeita H_0	Rejeita H_0
	$H_0: \mu_{FSI} + \mu_{FSE} = \mu_{MSI} + \mu_{MSE}$	Rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0
	$H_0: \mu_{FSI} = \mu_{FSE}$	Rejeita H_0	Não rejeita H_0	Não rejeita H_0	Não rejeita H_0

Nota: ¹média do grupo fêmeas no sistema intensivo; ²média do fêmeas no sistema extensivo; ³média do machos no sistema intensivo; ⁴média do machos no sistema extensivo.

Deve-se aplicar os testes, limitando a probabilidade de erro tipo *I* em um valor pré-atribuído, α , em geral entre 1% e 10%. Sob estas condições, espera-se minimizar a probabilidade

de erro tipo *II* (Baron, 2006). O poder do teste é a probabilidade de se rejeitar a hipótese nula H_0 , quando ela é realmente falsa e é dada por $1 - \beta$ (Mood et al., 1974).

As probabilidades de se cometerem os erros do tipo *I* e *II* são inversamente proporcionais, ou seja, a baixa probabilidade de se incorrer no erro tipo *I* está associada à alta probabilidade de se cometer o erro tipo *II* (Ferreira, 2005). Portanto, é necessário tomar algumas precauções para não se aplicar um teste a um nível muito baixo de probabilidade, para não aumentar exageradamente a probabilidade de ocorrência do outro.

Segundo Ferreira (2005), existem algumas medidas que podem ser tomadas para minimizar o erro tipo *II*. Uma que pode ser destacada consiste na escolha apropriada do teste e a avaliação cautelosa das pressuposições a serem atendidas. Se essas condições são satisfeitas, é provável que o poder do teste seja maior. Outra medida que pode permitir que um teste tenha o maior poder possível é a determinação do tamanho da amostra, desde que o custo da pesquisa a ser realizada não aumente excessivamente. Finalmente, a fixação do nível de significância α entre 0,10 e 0,01 é outra alternativa recomendável, por causa da relação inversa entre as taxas de erro tipo *I* e do erro tipo *II*.

A análise de variância (teste *F*) tem a finalidade de comparar estimativas de variâncias obtidas pelos quadrados médios para cada causa da variação. Essa primeira análise frequentista foi baseada nas estimações dos parâmetros envolvendo a soma de quadrados individuais do modelo e posteriormente, as estimações tornaram-se variáveis respostas em uma análise de variância. O teste de Tukey foi usado para comparação das diferenças entre médias, que tem pouca robustez. Na variável k_1 fez-se o uso de transformações Box-Cox (1964) para atender as pressupostos de normalidade e homogeneidade, apesar de alguns autores criticarem o uso de transformações pela falta de interpretação biológica, uma vez que pode ocorrer uma distorção dos valores médios quando realiza-se a transformação inversa.

A análise frequentista baseada na soma de quadrados da amostra total de cada grupo, realizada pelo teste *t*-student, apesar de apresentar uma boa convergência pontual, em relação as demais metodologias apresentou erros padrão elevados na maioria dos casos. Essa análise sofre forte influência quanto ao tamanho da amostra pois é baseada no conceito de grau de liberdades, quando o tamanho da amostra é pequeno ($n < 30$) pode produzir resultados não condizentes com a realidade, como no caso com menor precisão, embora mostrou-se acurado.

Os resultados *bootstrap* ajudam a melhorar a precisão e a acurácia com as réplicas de reamostragens geradas a partir de uma amostra que são tratadas como se fosse uma população. Sob certas condições de regularidades, o método *bootstrap* se aproxima da distribuição amostral do estimador ou da estatística. Segundo Shao e Tu (1995) e Horowitz (2001), essa aproximação assintótica apresenta uma precisão melhor para casos multivariados. O conceito de grau de liberdade ainda é usado na execução de cada solução de reamostragem. O estudo *bootstrap* melhora a acurácia do modelo obtendo resultados expressivos e pode ser uma análise preliminar de quando nada se conhece a respeito da distribuição dos parâmetros.

Os resultados das simulações Monte Carlo foram baseadas nas estimativas dos parâmetros das distribuições de densidades encontradas pela metodologia *bootstrap* não-paramétrico. As simulações de Monte Carlo suavizaram as densidades das distribuições dos parâmetros e a medida que o número de simulações aumentam espera-se aproximar à densidade teórica da distribuição normal que foi atribuída para cada cenário. Todos os parâmetros são vistos como variáveis aleatórias contemplando a incerteza existente sobre todos os parâmetros simultaneamente e a obtenção dos intervalos de confiança. A geração de sequências aleatórias é certamente um fator primordial a ser considerado, para precisão das estimativas Monte Carlo. O suporte estatístico para a validação dos resultados obtidos por simulação Monte Carlo é verificado na “lei dos grande números”, permitindo interpretar que à medida que o número de simulações aumenta, a estimativa converge para o valor verdadeiro do parâmetro. A utilização dessa metodologia requer um número menor de observações, pois os conceitos probabilísticos envolvidos diminuem a dependência do ajuste do modelo em relação ao número de observações utilizadas, uma vez que o conceito de graus de liberdade não é utilizado.

4.4 Conclusões

A técnica de *bootstrap* não-paramétrico com método de otimização *simplex downhill* estimou com sucesso as distribuições dos parâmetros (que tiveram distribuição de probabilidade normal para maioria dos cenários). As distribuições de densidades obtidas por Monte Carlo produziram curvas mais suavizadas em relação aos resultados obtidos pelo *bootstrap*, resultados de ambas metodologias apresentaram ajustes semelhantes para o modelo de Crescimento de Davis. Correlação negativa entre os parâmetros de acréscimo de DNA e energia de manutenção foram encontrados para animais machos não castrados do sistema extensivo, indicando que foram mais eficientes no uso da energia. A generalização de tal relação ainda demanda estudos mais abrangentes e aprofundados. Intervalos de confiança construídos por meio das análises frequentistas, *bootstrap* não-paramétrico e simulações Monte Carlo permitiram inferir que os parâmetros do modelo são diferentes dos originais determinados em Oltjen et al. (1986). Raças e melhoramento genético são as causas prováveis de alterações dos parâmetros. Os genótipos cruzados tiveram maior gasto de energia de manutenção e foram mais rápidos de maturação em comparação tanto com os animais de genótipos Britânicos (*Bos Taurus*) do estudo do Oltjen et al. (1986) como o estudo para touros Nelores realizados por Sainz et al. (2006). As novas estimativas de valores de parâmetro, em vez dos originais, podem ser usadas para fins preditivos para Nelore e mestiços.

4.5 Trabalhos futuros

Como continuidade desse estudo, seria interessante que trabalhos futuros tratassem dos seguintes assuntos:

- Avaliar o modelo considerando um banco de dados independente;

- Estimar a ingestão de matéria seca usando o MCD com novos banco de dados e comparar com outros modelos de estimação como Cornell Cattle Value Discovery System (CVDS), esse último modelo é amplamente utilizado em operações de confinamento comerciais nos Estados Unidos, constituído do National Research Council (NRC, 2000).

Agradecimentos

O presente trabalho foi realizado com o apoio financeiro da CAPES/Embrapa. A Embrapa Gado de Corte e Laboratório Crescimento de Nutrição Animal ESALQ/USP pelo fornecimento dos dados e dos subsídios, bem como a Embrapa Amazônia Oriental e a Embrapa Informática Agropecuária pelo grupo de parceria científica da rede Pecus e AnimalChange.

Referências

- BARON, M. **Probability and statistical for computer scientists**. New York: Chapman & Hall, 2006. 413 p.
- BELUM, V. R.; ROSEN, A. R.; JAIMES, N.; DRANITSARIS, G.; PULITZER, M. P.; BUSAM, K. J.; MARGHOOB, A. A.; CARVAJAL, R. D.; CHAPMAN, P. B.; LACOUTURE, M. E. Clinico-morphological Features of BRAF Inhibition-Induced Proliferative Skin Lesions in Cancer Patients. **Cancer**, Bethesda, v. 121, n. 1, p. 60-68, Jan. 2015.
- BHOWMICK, A.R.; BHATTACHARYA, S. A new growth curve model for biological growth: Some inferential studies on the growth of *Cirrhinus mrigala*. **Mathematical Biosciences**, Kolkata, v. 254, p. 28-41, 2014.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society**. London, v. 26, n. 1, p. 211-252, Abr. 1964.
- CHIZZOTTI, M.L.; TEDESCHI, L.O.; VALADARES FILHO, S.C. A meta-analysis of energy and protein requirements for maintenance and growth of Nellore cattle. **Journal of Animal Science**, Champaign, v. 86, p. 1588-1597, 2008.
- DIEBOLD, F.X.; OHANIAN, L.E.; BERKOWITZ, J. Dynamic Equilibrium Economies: A Framework for Comparing Models and Data. **Review of Economic Studies**, v. 65, p. 433-451, 1998.
- DRAPER, N.R.; SMITH, H. **Applied regression analysis**. 3.ed. New York: John Wiley e Sons, 1998. 706p.
- FERREIRA, D. F. **Estatística Básica** Lavras: UFLA, 2005. 664 p.
- FREETLY, H.C.; KUEHN, L.A.; CUNDIFF, L.V. Growth curves of crossbred cows sired by Hereford, Angus, Belgian Blue, Brahman, Boran, and Tuli bulls, and the fraction of mature body weight and height at puberty. **Journal Animal Science**, Champaign, v. 89, p. 2373-2379, 2011.

- FRANCE, J.; GILL, M.; THORNLEY, J.H. M.; ENGLAND, P. A model of nutrient utilization and body composition in beef cattle. **Animal Production**, England, v. 44, 371-385, 1987.
- FORNI, S.; PILES, M.; BLASCO, A.; VARONA, L.; OLIVEIRA, H.N.; LÔBO, R.B.; ALBUQUERQUE, L.G. 2009. Comparison of different nonlinear functions to describe Nelore cattle growth. **Journal Animal Science**, Champaign, v. 87, p. 496-506.
- HOCH, T.; AGABRIEL, J. A mechanistic dynamic model to estimate beef cattle growth and body composition:1. Model description. **Essex**, United Kingdom, n. 81, p. 1-15, 2004.
- HOGG, R.; TANIS, E. **Probability and Statistical Inference**. 8. ed. New Jersey: Person Prentice Hall, 2010. 622 p.
- HOROWITZ, J.L. The Bootstrap. In: HECKMAN, J.J; LEAMER, E.E. (Eds.), North Holland: Elsevier Science, Handbook of Econometrics, n. 5, 2001, p. 3159-3228.
- LI, Z.; SHAO, Q.; XU, Z.; CAI, X. Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: A case study of SWAT model applied to Yingluoxia watershed in northwest China. **Journal of Hydrology**. v. 385, n. 1, p. 76-83, Jan. 2010.
- MARCONDES, M.I.; CHIZZOTTI, M.L.; VALADARES FILHO, S.C.; GIONBELLI, M.P.; PAULINO, P.V.R.; PAULINO, M.F. Energy requirements of zebu beef cattle. p. 81-106, 2010a. In: VALADARES FILHO, S.C.; MARCONDES, M.I.; CHIZZOTTI, M.L.; PAULINO, P.V. (Eds.) **Nutrient requirements of zebu cattle BR-CORTE**. 2.ed. DZO-UFV, Viçosa, MG, Brazil.
- MARCONDES, M.I.; CHIZZOTTI, M.L.; VALADARES FILHO, S.C.; GIONBELLI, M.P.; PAULINO, P.V.R.; PAULINO, M. F. Protein requirements of zebu beef cattle. p. 107-126, 2010b. In: VALADARES FILHO, S.C.; MARCONDES, M.I.; CHIZZOTTI, M.L.; PAULINO, P.V. (Eds.) **Nutrient requirements of zebu cattle BR-CORTE**. 2.ed. DZO-UFV, Viçosa, MG, Brazil.
- MARINHO, K.N.S.; FREITAS, A.R.; FALÇÃO, A.J.S.; DIAS, F.E.F. Nonlinear models for fitting growth curves of Nelore cows reared in the Amazon Biome. **Revista Brasileira de Zootecnia**, v. 42, p. 645-650, 2013.
- MATTHEWS, K.; ZHANG, N. Bank productivity in China 1997-2007: Measurement and convergence. **China Economic Review**, Pequim, v. 21, n. 1, p. 617-628, Jun. 2010.
- MCPHEE, M. J. et al. Parameter Estimation of Fat Deposition Models in Beef Steers. **DNA**, New Zelandia, v. 4, p. 115-120, Dec. 2007.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3. ed. Singapore: McGraw-Hill, 1974. 564 p.
- MUNIZ, L.C.; VIU, M.A. O.; MAGNABOSCO, C.U.; LOPES, D.T. Modeling and simulation of agricultural. **Pubvet**, Londrina, v. 11, n.1, p. 1982-1263, 2007.

NELDER, J. A.; MEAD, R. A simplex algorithm for function minimization. **Computer Journal**, Cambridge, v.7, n.1, p. 308-313, 1965.

NESETRILOVÁ, H. Multiphasic growth models for cattle. **Jornal Animal Science**, Champaign, v. 8, p. 347-354, 2005.

NRC. **Nutrient Requirements of Beef Cattle**. 7ed. National Academies Press, Washington, DC, 1996, 234 p.

NRC. **Nutrient Requirements of Beef Cattle**. 7ed. National Academies Press, Washington, DC, 2000, 248 p.

OLTJEN, J.W.; PLEASANTS, A.B; SOBOLEVA, T.K. ODDY, V.H. Second-generation dynamic cattle growth and composition models. p. 197-209, 2000. In: MCNAMARA, J.P.; FRANCE, J.; BEEVER, D.E. **Modelling Nutrient Utilization in Farm Animals**. CAB International, London, LDN, United Kingdom.

OLTJEN, J.W.; BYWATER, A.C.; BALDWIN R.L.; GARRETT W.N. Development of a dynamic model of beef cattle growth and composition. **Journal Animal Science**, v. 62, p. 86-97, 1986.

PAULINO, P.V.R.; FONSECA, M.A.; HENRIQUES, L.T.; VALADARES FILHO, S.C.; DETMANN, E. Nutritional Requirements of nellore cows and calves. p. 167-185, 2010. In: VALADARES FILHO, S.C.; MARCONDES, M.I.; CHIZZOTTI, M.L. et al. (Eds.) **Nutrient requirements of zebu cattle BR-CORTE**. 2.ed. DZO-UFV, Viçosa, MG, Brazil.

PRESS, W.H; FLANNERY, B.P; TEUKOLSKY, S.A; VETTERLING, W.T. **Numerical recipes in C: the art of scientific computing**. 2 ed. Cambridge University Press, NY, USA, 1990, 925 p.

PRESS, W.H.; TEUKOLSKY, S.A; VETTERLING, W.T.; FLANNERY B.P. **Numerical recipes in FORTRAN: the art of scientific computing**. 2 ed. Cambridge University Press, NY, USA, 1992, 554 p.

PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K. CODA: Convergence Diagnosis and Output Analysis for MCMC, **R News**, v. 6, n. 1, p. 7-11, Mar. 2006.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria, 2014. Disponível em: <<http://www.R-project.org>>. Acesso em: 20 mar. 2014.

SHAO, J.; TU, D. **The Jackknife and bootstrap**. New York, Springer, Verlag, 1995. 516 p.

SAINZ, R.D; BARIONI, LG; PAULINO; P.V.; VALADARES FILHO, S.C.; OLTJEN, J.W. Growth patterns of Nellore vs British Beef Cattle Breeds assessed using a dynamic, mechanistic model of cattle growth and composition. p. 160-170, 2006. In: KEBREAB, E.; DIJKSTRA, J.; BANNINK, A.; GERRITS, W.J.J.; FRANCE, J. **Nutrient Digestion and utilization in Farm Animals: modeling approaches**. CAB International, London, LDN, United Kingdom.

SOETAERT, K.; PETZOLDT, T. Inverse Modelling, Sensitivity and Monte Carlo Analysis in R Using Package FME. **Journal of Statistical Software**, Los Angeles, v.33, n.3, p. 1-28, Feb. 2010.

SOETAERT, K.; PETZOLDT, T.; SETZER, R. W. Solving Differential Equations in R: Package deSolve. **Journal of Statistical Software**, Los Angeles, v. 33, n.9, p. 1-25, Dec. 2010.

SRINIVAS, V.V.; SRINIVASAN, K. Matched block bootstrap for resampling multiseason hydrologic time series. **Hydrological Processes**. Bangalore, v. 19, n. 1, p. 3659-3682, May 2005.

STEEL, R. G. D.; TORRIE, J. H. DICKEY, D. A. **Principles and procedures of statistics: a biometrical approach**. 3. ed. New York: McGraw-Hill, 1997. 666 p.

TEDESCHI, L.O.; FOX, D.G.; GUIROY, P.J. A decision support system to improve individual cattle management. 1. A mechanistic, dynamic model for animal growth. **Agricultural Systems**, Barking, v. 79, p. 171-204, 2004.

VALENTE, E.E. L.; PAULINO, M.F.; DETMANN, E.; VALADARES FILHO, S.C.; CARDENAS, J.E.G.; DIAS, I.F.T. Requirement of energy and protein of beef cattle on tropical pasture. **Acta Scientiarum**, Maringá, v. 35, p. 417-424, 2013.

WALMSLEY, B.J.; WOLCOTT, M.L.; MCPHEE, M.J. Modeling the relationship between scanned rump and 12th-rib fat in young temperate and tropical bovines: Model development and evaluation. **Journal Animal Science**, Champaign, v. 88, n.1, p. 1848-1859, Dec. 2014

WILLIAMS, C.B.; BENNETT, G.L. Application of a computer model to predict optimum slaughter end points for different biological types of feeder cattle. **Journal Animal Science**, Champaign, v. 73, p. 2903-2915, 1995.

WILLIAMS, C.B.; JENKINS T.G. A dynamic model of metabolizable energy utilization in growing and mature cattle. II. Metabolizable energy utilization for gain. **Journal Animal Science**, Champaign, v. 81, p. 1382-1389, 2003.

ANEXO

Axexo A: Gráficos do Capítulo 2 - banco de dados sintético

Modelo analítico exponencial

Figura 4.5 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$ e $p3 = \theta_3$) estimados por MCMC, modelo analítico exponencial com perturbação $N(0; 0, 05^2)$

Figura 4.6 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 , θ_3) e variância (σ^2) do modelo analítico exponencial com perturbação $N(0; 0, 05^2)$

Figura 4.7 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$ e $p3 = \theta_3$) estimados por MCMC, modelo exponencial com perturbação $N(0; 0, 2^2)$

Figura 4.8 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) do modelo analítico exponencial com perturbação $N(0; 0, 2^2)$

Modelo analítico logístico

Figura 4.9 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$ e $p3 = \theta_3$) estimados por MCMC, modelo analítico logístico com perturbação $N(0; 0, 05^2)$

Figura 4.10 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) do modelo analítico logístico com perturbação $N(0; 0, 05^2)$

Figura 4.11 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$ e $p3 = \theta_3$) estimados por MCMC, modelo logístico com perturbação $N(0; 0, 2^2)$

Figura 4.12 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) do modelo analítico logístico com perturbação $N(0; 0, 2^2)$

Modelo analítico logístico dual-pool

Figura 4.13 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1, p2 = \theta_2, p3 = \theta_3, p4 = \beta_4$ e $p5 = \beta_5$) estimados por MCMC, modelo analítico logístico bicompartimental com perturbação $N(0; 0, 05^2)$

Figura 4.14 – Traços das cadeias e distribuições de probabilidade dos parâmetros ($\theta_1, \theta_2, \theta_3, \theta_4$ e θ_5) e variância (σ^2) do modelo analítico logístico bicompartimental com perturbação $N(0, 0, 05^2)$

Figura 4.15 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($p1 = \theta_1$, $p2 = \theta_2$, $p3 = \theta_3$, $p4 = \beta_4$ e $p5 = \beta_5$) estimados por MCMC, modelo analítico logístico bicompartimental com perturbação $N(0; 0, 2^2)$

Figura 4.16 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 , θ_3 , θ_4 e θ_5) e variância (σ^2) do modelo analítico logístico bicompartimental com perturbação $N(0; 0, 2^2)$

Equação diferencial ordinária exponencial

Figura 4.17 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($a = \theta_1$, $b = \theta_2$ e $c = \theta_3$) estimados da EDO exponencial, MCMC, com perturbação $N(0; 0, 05^2)$

Figura 4.18 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) com EDO exponencial com perturbação $N(0; 0, 05^2)$

Figura 4.19 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($a = \theta_1$, $b = \theta_2$ e $c = \theta_3$) estimados por MCMC, com EDO exponencial com perturbação $N(0; 0, 2^2)$

Figura 4.20 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) do modelo analítico exponencial com perturbação $N(0; 0, 2^2)$

Equação diferencial ordinária logístico

Figura 4.21 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($a = \theta_1$, $b = \theta_2$ e $c = \theta_3$) estimados da EDO logístico, MCMC, com perturbação $N(0; 0, 05^2)$

Figura 4.22 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) com EDO logístico com perturbação $N(0; 0, 05^2)$

Figura 4.23 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($a = \theta_1$, $b = \theta_2$ e $c = \theta_3$) estimados da EDO logístico, MCMC, com perturbação $N(0; 0, 2^2)$

Figura 4.24 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 e θ_3) e variância (σ^2) com EDO logístico com perturbação $N(0; 0, 2^2)$

Equação diferencial ordinária logística bicompartimental

Figura 4.25 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros ($a = \theta_1$, $b = \theta_2$, $c = \theta_3$, $d = \theta_4$ e $e = \theta_5$) estimados da EDO logístico bicompartimental, MCMC, com perturbação $N(0; 0, 05^2)$

Figura 4.26 – Traços das cadeias e distribuições de probabilidade dos parâmetros (θ_1 , θ_2 , θ_3 , θ_4 e θ_5) e variância (σ^2) com EDO logístico bicompartimental com perturbação $N(0; 0, 05^2)$

Figura 4.27 – Matriz de dispersão, histograma de frequência e correlações dos parâmetros estimados da EDO logístico bicompartimental, MCMC, com perturbação $N(0; 0, 2^2)$

Figura 4.28 – Traços das cadeias e distribuições de probabilidade dos parâmetros ($\theta_1, \theta_2, \theta_3, \theta_4$ e θ_5) e variância (σ^2) com EDO logístico bicompartimental com perturbação $N(0; 0, 2^2)$