CARLOS ANTÔNIO ZARZAR

ESTATÍSTICA APLICADA À AQUICULTURA, DESAFIOS E MÉTODOS PARA A MODELAGEM DE CRESCIMENTO DO CAMARÃO CINZA

LAVRAS – MG

2023

CARLOS ANTÔNIO ZARZAR

ESTATÍSTICA APLICADA À AQUICULTURA, DESAFIOS E MÉTODOS PARA A MODELAGEM DE CRESCIMENTO DO CAMARÃO CINZA:

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, na área de Ciências Agrárias, para a obtenção do título de Doutor.

Prof. Dra. Izabela Regina Cardoso de Oliveira
Orientadora

Prof. Dr. Tales Jesus Fernandes
Coorientador

LAVRAS – MG
2023

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Zarzar, Carlos Antônio.
Estatística aplicada à aquicultura, desafios e métodos para a modelagem de crescimento do camarão cinza / Carlos Antônio Zarzar. - 2023.
97 p. : il.

Orientador(a): Izabela Regina Cardoso de Oliveira.
Coorientador(a): Tales Jesus Fernandes.
Tese (doutorado) - Universidade Federal de Lavras, 2023.
Bibliografia.

1. Aquicultura 40. 2. Carcinicultura. 3. Modelos Bayesianos. I. de Oliveira, Izabela Regina Cardoso. II. Fernandes, Tales Jesus. III. Título.

CARLOS ANTÔNIO ZARZAR

ESTATÍSTICA APLICADA À AQUICULTURA, DESAFIOS E MÉTODOS PARA A MODELAGEM DE CRESCIMENTO DO CAMARÃO CINZA
STATISTICS APPLIED TO AQUACULTURE, CHALLENGES AND METHODS FOR MODELING PACIFIC WHITE SHRIMP GROWTH

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, na área de Ciências Agrárias, para a obtenção do título de Doutor.

APROVADA em 23 de Junho de 2023.

Prof. Dr. Julio Silvio de Sousa Bueno Filho	UFLA
Prof. Dr. Luis Otavio Brito da Silva	UFRPE
Prof. Dr. Cristian Marcelo Villegas Lobos	ESALQ/USP

Prof. Dra. Izabela Regina Cardoso de Oliveira
Orientadora

Prof. Dr. Tales Jesus Fernandes
Co-Orientador

LAVRAS – MG
2023

Dedico essa tese a todos familiares, companheira, amigos, orientadores e todos que fizeram parte na trajetória desse doutorado.

AGRADECIMENTOS

Um grupo de especialistas em Ciências de Aprendizagem estudaram a habilidade e competência de vários alunos para entender o motivo de que alguns aprendem mais rápido do que outros. A descoberta dessa investigação foi surpreendente. Segundo os autores a taxa de aprendizagem é igual para todos, o motivo de uns aprenderem mais rápido do que outros depende das oportunidades que o indivíduo teve nos estudos e a aplicação desse conhecimento. Portanto, nas condições adequadas, as pessoas aprendem em um ritmo notavelmente semelhante.

Essa estranha introdução em um agradecimento é apenas para enfatizar a importância das oportunidades na nossa trajetória reconhecendo a educação e o esforço da minha mãe Renilda M. N. Aranha ao longo da minha vida. Foram essas raízes familiares (mãe, pai, meus irmãos Sáfia e Toninho, tios, avôs e avós) que motivaram sobrepujar todas as dificuldades, consequentemente formou minha personalidade e me motivou para materializar as minhas oportunidades.

Agradecimento especial a minha companheira Raphaela L. de M. Silva que também construiu essa tese comigo, me apoiando, me acompanhando em todas as aventuras e imaginação que um bom geminiano gosta de se jogar. Como estávamos os dois trabalhando em home office (a partir da pandemia) compartilhávamos agradavelmente a mesa da sala um de frente para o outro, entre computadores e monitores todos os dias durante os 2 últimos anos do doutorado.

Agradecimentos a minha orientadora Izabela R. C. Oliveira, coorientador Tales J. Fernandes, aos professores participantes da banca, ao programa da Pós-graduação da UFLA, grupos de estudos, amigos e colaboradores que fizeram parte dessa história. E principalmente a Universidade Federal do Oeste do Pará (UFOPA) Campus Monte Alegre, amigos de trabalho técnicos, alunos e professores, o qual eu e meus companheiros dedicamos nossas vidas a transformar a região.

Agradeço ao senhor Eduardo de Freitas D. Antona que através da minha irmã Sáfia A. Zarzar proporcionaram a colaboração e início dessa pesquisa científica e consequentemente nessa tese.

E não menos importante agradeço a Deus e todos os espíritos que envolvem a nossa evolução na terra e em outros planos superiores. Gratidão a todos.

E como todo bom pesquisador faz, nunca deixamos de referenciar nossos argumentos, segue a referência do artigo citado sobre ciências de aprendizagem para interessados (até no agradecimento temos citações? Só para descontrair com humor científico): KOEDINGER, Kenneth R. et al. An astonishing regularity in student learning rate. Proceedings of the National Academy of Sciences, v. 120, n. 13, p. e2221311120, 2023.

O que diferencia um gênio de uma pessoa extremamente inteligente é uma única coisa, a criatividade. Exorte sua imaginação.
(Carlos A. Zarzar)

RESUMO

A aquicultura (cultivo de organismos aquáticos) no Brasil vem se destacando no agronegócio nos últimos anos, apesar de ser um setor relativamente recente quando comparado com a bovinocultura e avicultura. A produção aquícola do Brasil em 2020 segundo o IBGE foi de 629,3 mil toneladas. A vasta rede hidrográfica (75 mil km) em rios, lagos e lagoas (cerca de 167 mil km^2), além da extensão da costa litorânea (7.367 km do Amapá ao Rio Grande do Sul), aliado com o clima favorável, tornam o Brasil um país com grande potencial para o setor. Para destacar a aquicultura frente o mercado competitivo (nacional e internacional) é necessário acompanhar a evolução dos novos recursos estatísticos e da inteligência artificial na busca de produções mais eficientes. Dentro do universo empresarial aquícola a receita, os custos e lucros são baseados no peso da proteína animal vendida no mercado. Portanto a modelagem de crescimento de organismos cultivados é utilizada como ferramenta de gestão da produção. Alguns dados de crescimento na aquicultura possuem características peculiares que geram consequências na análise e modelagem. Geralmente são incompletos ou limitados. Isso significa que os dados são restritos a poucas observações e muitas vezes são limitados a observações abaixo do ponto de inflexão da curva sigmoide devido a estratégia econômica das fazendas ou simplesmente a exigência do mercado consumidor. Essa limitação dos dados observados presumivelmente causa viés na inferência de modelos não lineares. Por meio de simulações de crescimento do camarão com dados limitados e comparações de curvas de crescimento de animais selvagens oriundos da pesca, os resultados apoiaram essa hipótese. Como consequência, foi proposto um método para correção deste possível viés por meio de uma abordagem bayesiana hierárquica. Dados reais foram utilizados para compará-la com a abordagem frequentista tradicional utilizada. A sensibilidade em detectar o melhor tratamento pode fazer com que o novo método seja uma poderosa ferramenta de gerenciamento na produção animal, inclusive em ensaios delineados para pesquisa científica. No segundo capítulo, com base na metodologia bayesiana proposta, seis modelos hierárquicos não lineares foram avaliados para modelagem do crescimento do camarão cinza (*Litopenaeus vannamei*) (Morgan-Mercer-Flodin, Michaelis-Menten, Weibull, von Bertalanffy, Gompertz e Logístico) e ajustados aos dados reais de uma fazenda de produção. A equação de crescimento de Weibull se destacou. O modelo final foi validado mostrando uma precisão de 95,76% e 85,71% nos níveis hierárquicos de viveiro e ciclo produtivo, respectivamente. Finalmente, uma análise de sensibilidade foi realizada para detectar diferenças sutis entre cultivos aparentemente semelhantes e concluímos que a nova abordagem é muito eficiente para comparação de tratamentos quando contraposto a índices zootécnicos usualmente praticados nas fazendas comerciais de camarão.

Palavras-chave: Aquicultura 4.0. Carcinicultura. Modelos Bayesianos. Ciência de Dados. Algoritmo

ABSTRACT

Aquaculture in Brazil has been standing out in agribusiness in recent years, despite being a relatively recent sector when compared to cattle and poultry. Aquaculture production in Brazil in 2020 was 629.3 thousand tons. The vast hydrographic network (75,000 km) in rivers, lakes, and ponds (about 167,000 km^2), in addition to the extension of the coastline (7,367 km from Amapá to Rio Grande do Sul), combined with the favorable climate, given Brazil a country with great potential for the sector. To highlight aquaculture, in a competitive market (national and international), it is necessary to keep up the evolution of statistical methods and artificial intelligence in the search for more efficient production. Within the aquaculture business, revenue, costs, and profits are based on the animal weight. Therefore, the growth modeling of organisms is used as a production management tool. Some growth data in aquaculture have peculiar characteristics that generate consequences for analysis and modeling. They are usually incomplete or limited. This means that the data are restricted to a few observations and are often limited to observations below the inflection point of the sigmoid curve. This occurs due to the economic strategy of the farms or simply the demand of the consumer market. This limitation of the observed data presumably causes bias in the inference of nonlinear models. Results from simulations and comparisons between the growth of wild animals and fisheries supported this hypothesis. As a result, a method was proposed to correct this possible bias using a hierarchical Bayesian approach. Real data were used to compare it with the traditional frequentist approach used. The sensitivity in detecting the best treatment can make the new method a powerful management tool in animal production, including trials designed for scientific research. In the second chapter, based on the proposed Bayesian methodology, six non-linear hierarchical models were evaluated for modeling the growth of gray shrimp (*Litopenaeus vannamei*) (Morgan-Mercer-Flodin, Michaelis-Menten, Weibull, von Bertalanffy, Gompertz, and Logistics) and adjusted to real data from a production farm. The Weibull growth equation stood out. The final model was validated showing an accuracy of 95.76% and 85.71% in the hierarchical pond and production cycle levels, respectively. Finally, a sensitivity analysis was carried out to detect subtle differences between crops and we concluded that the new approach is very efficient for comparing treatments.

Keywords: Aquaculture 4.0. Shrimp farming. Bayesian Models. Data Science. Algorithm.

LISTA DE FIGURAS

SUMÁRIO

PRIMEIRA PARTE

1 INTRODUÇÃO

A aquicultura pode ser definida como o cultivo de organismos aquáticos tais como peixes (piscicultura), camarões (carcinicultura), algas (algicultura), rãs (ranicultura), jacaré (jacarecultura) ou qualquer outro animal ou vegetal que estejam relacionados aos ambientes marinhos ou dulcícolas (água doce). É uma atividade praticada há milhares de anos, remontando às antigas civilizações do Egito, China e Roma. No entanto, a aquicultura moderna como a conhecemos hoje tem suas raízes no século XIX (NASH, 2010).

Nas últimas décadas a aquicultura mundial vem crescendo e se destacando como uma atividade competitiva sustentável na produção de alimentos saudáveis, apresentando contribuição relevante para geração de emprego e renda, bem como redução da pobreza e da fome em várias partes do mundo. De acordo com as estatísticas mundiais sobre aquicultura compiladas pela FAO (2020), a produção mundial de animais aquáticos de criação cresceu em média 5,3% ao ano no período 2001-2018. A produção mundial atingiu outro recorde histórico de 114,5 milhões de toneladas de peso vivo em 2018 (correspondendo a uma economia US$ 263,6 bilhões) com projeções que superou a produção extrativista da pesca em 2020. No cenário nacional não é diferente. A produção do Brasil em 2018 foi de 605 mil toneladas e segundo as projeções da FAO (2020) de 2018 até 2030 ocorrerá um significativo crescimento de 32,2% para o setor produtivo nacional.

Devido a importância dessa atividade econômica para o Brasil e sua lucratividade, esse setor constantemente atrai investidores diretos (empreendimentos de fazendas de produção) ou indiretos (empresas de consultoria ou tecnologias) que buscam maximizar a eficiência no cultivo e minimizar seus custos. A união da aquicultura com a estatística e a ciência de dados formam a perfeita sinergia na consecução desse objetivo que infelizmente ainda é pouco explorado. Ao longo da tese foi apresentado tecnologias através de métodos estatísticos que podem melhorar o setor aquícola e aumento da eficiência na produção. Portanto, o objetivo principal é a modelagem do crescimento do camarão cinza (*Litopenaeus vannamei*) considerando dados reais de um ambiente produtivo de uma fazenda do nordeste brasileiro.

O projeto teve parceria público-privado, com a participação do Grupo de investidores SECOM que forneceu os dados de uma das suas fazendas de camarão localizada no litoral do estado do Piauí. Dessa forma, a maioria das ferramentas e métodos estatísticas desenvolvi-

dos foram aplicados a carcinicultura para a espécie de camarão cinza (*Litopenaeus vannamei*). No entanto, sua aplicação não se restringe a essa espécie ou até mesmo apenas ao crustáceo, podendo ser aplicados a qualquer produção de interesse econômico.

O primeiro capítulo levanta a hipótese, até então nunca relatado, de que os parâmetros dos modelos de crescimento não lineares (funções sigmoides) utilizados usualmente na aquicultura, podem estar subestimados. Em virtude das amostras do peso dos camarões serem restritos a observações abaixo do ponto de inflexão (no qual foram classificados como dados incompletos/limitados), podem levar a estimativas viesadas dos parâmetros em modelos não lineares de crescimento. Tais parâmetros muitas vezes possuem significados biológicos como o peso assintótico, taxa de crescimento ou até mesmo idade (tamanho) de maturação. Portanto estimativas não viesadas desses parâmetros são de grande importância para esta ciência aplicada. A investigação comparou as estimativas dos parâmetros em modelos na aquicultura com os modelos ajustados para a mesma espécie utilizando dados oriundos da pesca (dados sem a restrição ou a limitação). Além do mais, foram simulados dados incompletos de crescimento de camarão (incorporando a heterogeneidade do peso ao longo do tempo) em diferentes *threshold* de limitações da informação para quantificar a sensibilidade dessa possível subestimativa na modelagem em funções não lineares de crescimento corroborando com a hipótese da pesquisa. Por fim, no segundo momento da investigação foi proposto um método baseado em modelos Bayesianos Hierárquicos para correção e validação utilizando os dados reais da fazenda.

No segundo capítulo, com base na metodologia Bayesiana proposta para corrigir o viés de estimação dos parâmetros, seis modelos hierárquicos de crescimento não linear foram avaliados nessa pesquisa (Morgan-Mercer-Flodin, Michaelis-Menten, Weibull, von Bertalanffy, Gompertz e equação de crescimento Logístico) e ajustados a dados reais de uma fazenda de camarão cinza (*Litopenaeus vannamei*) no nordeste do Brasil. O modelo resultante dessa investigação foi selecionado, diagnosticado e validado. Uma análise de sensibilidade foi realizada para detectar diferentes crescimentos do camarão entre diferentes características ao longo dos cultivos da fazenda. De forma genérica a nova abordagem detectou diferenças sutis entre cultivos aparentemente semelhantes (baseado em índices zootécnicos tradicionais), que seriam imperceptíveis se analisadas por métodos comparativos usuais praticados nas empresas comerciais.

2 REFERENCIAL TEÓRICO

Todos os vetores definidos serão representados por quantidades multivariadas ao longo do texto, porém não serão escritos em negrito, seguindo o mesmo padrão da maioria das referências bibliográficas mencionadas. Em notações gerais denominaremos θ como vetor de parâmetros não observáveis de interesse o qual pertence ao conjunto do espaço paramétrico desconhecido denominado de Θ. Referente ao método numérico para o cálculo da função densidade de probabilidade a posteriori dos modelos Bayesianos, denota-se θ^S como o vetor de parâmetros iterativo obtido pelo algoritmo de integração numérica utilizado. y refere-se ao vetor de variáveis aleatórias, correspondente aos dados observados. $\tilde{y}$ refere-se a um vetor de quantidade desconhecida, mas potencialmente observável (uma estimativa ou uma predição da variável y, em oposição ao valor real observado). Usaremos letras maiúsculas para denotar variáveis aleatórias e letras minúsculas para denotar valores observados da amostra.

2.1 Camarão *Litopenaeus vannamei* e a carcinicultura

O camarão cinza (*Litopenaeus vannamei*), também conhecido como camarão branco do Pacífico (Figura 2.1), é uma das espécies de camarão mais consumida do mundo. É endêmico do Oceano Pacífico distribuído na costa Oeste das Américas, desde o México (na Província de Sonora) até o Peru (ao Sul de Tumbes) (BENZIE, 2000). No entanto, o cultivo em fazendas aquícolas ultrapassa qualquer limitação espacial restrita a pesca e sobrepuja as fronteiras físicas limitadas a populações selvagens dessa região. O cultivo de crustáceos (carcinicultura) possibilitou alcançar grandes mercados internacionais, produzido por mais de 50 países pertencentes principalmente a Ásia, América Central e América do Sul (CASTRO; CAVALCANTI-MATA; DUARTE, 2004).

O *Litopenaeus vannamei* é um crustáceo decápota, pertencente à família Penaeidae. É um camarão de água salgada que se desenvolve muito bem em salinidade entre 15 e 30%, em temperaturas de 23 a 30°C. Entretanto, por ser uma espécie eurihalina que tolera salinidades de 0,5 (ROY; DAVIS, 2010) até 60% (CHONG-ROBLES et al., 2014), é comum encontrar seu cultivo em baixas salinidades em regiões interiores de vários países como Tailândia, EUA e Brasil (BOYD; THUNJAI, 2003; DAVIS; SAMOCHA; BOYD, 2004; NUNES, 2001; ROY; DAVIS, 2010), que buscam novos mercados de camarão frescos distantes da costa. A sua rusti-

Figura 2.1 – Figura do camarão cinza (*Litpenaeus vannamei*) também conhecido como camarão Branco do Pacífico

Fonte: repositório gratuito *unsplash* disponibilizado por James Tiono (2022)

cidade, adaptabilidade e crescimento rápido (pode atingir um tamanho comercial em 120 dias) são as principais características que chamam a atenção do setor aquícola. A carne do animal tem uma coloração típica e boa qualidade nutricional de excelente aceitação em mercados internacionais (SCHOBER, 2002), além da tolerância de altas densidades de estocagem e consumo do alimento natural presentes no tanque de cultivo (SAMOCHA et al., 2017) o que torna uma atividade econômica muito rentável.

Acredita-se que o cultivo do camarão marinho teve origem histórica no Sudoeste da Ásia, onde pescadores artesanais construíam diques de terra nas zonas costeiras para a captura de pós-larvas selvagens que cresciam nas condições naturais. O início da atividade no Brasil data na década de 70 no Rio Grande do Norte (RN), inicialmente proposto como alternativa a tradicional produção de sal do estado (NUNES, 2001; BARBOSA, 2022) e em Santa Catarina (SC) com as primeiras reproduções em cativeiro na produção de pós-larva em laboratório do Brasil (NUNES, 2003).

Desde então a carcinicultura (Figura 2.2) é uma atividade sólida no país apresentando um rápido crescimento na década de 90 até seu ápice em 2003 na região do Nordeste. A produção nacional foi de 63,2 mil toneladas em 2020, gerando economia para o país com um Valor Bruto de Produção de USD 282,14 milhões de dólares em 2020 (IBGE, 2020), e crescimento substancial de $23,8\%$ em 2021. Além de representar uma importante alternativa para a demanda mundial por proteína animal de boa qualidade, a atividade sócio-econômica do setor favorece as regiões rurais produtivas evitando o êxodo rural. Portanto, é natural o crescimento científico e os esforços em diversas áreas de interesse para atender a crescente demanda por esse alimento em todo o mundo.

Figura 2.2 – Fazenda representando a cacinicultura

Fonte: *Wikimedia Commons* licença direitos autorais livre para compartilhar (1984)

2.2 Estatística Bayesiana

Os princípios Bayesianos se diferenciam da estatística clássica desde sua concepção filosófica até a inferência sobre a população em estudo. Alguns afirmam que os métodos Bayesianos se passam, de certa forma, por uma extensão do modelo clássico. Porém, divergem sobre a fundamental percepção dada ao parâmetro de um modelo de probabilidade de interesse.

A probabilidade é um conceito abstrato fundamentado por teoremas e introduzido por axiomas, em que se estudam as mais importantes propriedades. Entretanto, no campo abstrato interpretativo está susceptível a diversas perspectivas como foi caracterizado por Paulino, Turkman e Murteira (2003):

a) **A probabilidade como uma interpretação clássica** baseada em resultados igualmente possíveis ou prováveis, pode ser definida como um cociente entre o número de casos favoráveis e o número de casos possíveis de um evento ocorrer, supondo que todos os casos são equiprováveis;

b) **A interpretação frequentista** que se baseia na regularidade estatística das frequências relativas, definindo que a probabilidade de um dado acontecimento pode ser medida observando a frequência relativa do mesmo acontecimento numa sucessão numerosa de provas ou experimentos, idênticas e independentes;

c) **A interpretação lógica** ou “necessária” defende que a probabilidade representa uma relação lógica entre uma proposição (a evidência) e outra proposição (a hipótese) medindo assim, o grau de implicação da hipótese pela evidência. Sendo que esse grau de implicação é único, racional e impessoal. Porém, como essa interpretação era de difícil medição, tornou-se pouco operacional. No entanto, a sua evolução conceitual levou a possibilidade operacional com a decorrente interpretação;

d) **A interpretação subjetiva** ou personalista, considera que a probabilidade também é uma relação entre a evidência e a hipótese dos fatos, porém medindo o grau de credibilidade que "uma dada pessoa", à luz da evidência, atribui à hipótese. Essa interpretação subjetiva levou a estatística Bayesiana, fundamentada no Teorema de Bayes (GELMAN et al., 2013), a controversas com a estatística clássica;

A estatística clássica se baseia na interpretação das conclusões à luz da amostragem. Portanto, sua inferência consiste em reconhecer a variabilidade que se verifica na amostra e que os dados observados foram apenas um dos muitos (possivelmente infinitos) conjuntos que poderiam ter sido obtidos nas mesmas circunstâncias. Tal perspectiva leva a interpretação dos dados em função não apenas do particular conjunto observado, mas também das hipóteses adotadas acerca dos possíveis conjuntos alternativos de dados.

Dessa forma, induz a aceitar os dados como observação de uma variável aleatória Y ou de n variáveis aleatórias $Y = (Y_1, Y_2, \ldots, Y_n)$ com função de distribuição F_0 que representa a variabilidade ou a incerteza na observação da variável Y. Evidentemente que F_0 não é perfeitamente conhecida. No entanto, normalmente existe algum conhecimento inicial sobre a natureza do fenômeno aleatório em estudo ou sobre o processo gerador dos dados que leva a propor uma família de distribuições $\mathscr{F}$ a que pertence F_0 e que se designa por modelo estatístico para Y. A proposta de um modelo é essencial para a inferência clássica. Se as distribuições de $\mathscr{F}$ são representadas pelas respetivas densidades (função densidade de probabilidade ou função de probabilidade) e estas forem governadas por um parâmetro θ de dimensão p com domínio em um conjunto Θ, designado espaço-paramétrico, o modelo estatístico pode ser escrito:

$$\mathscr{F} = \{f(y \mid \theta) \ : \theta \in \Theta, \ y \in Y\}.$$

Portanto, na perspectiva clássica o parâmetro θ pode ser um escalar ou um vetor desconhecido, porém fixo. No modelo Bayesiano o parâmetro θ é tomado como um escalar ou vetor aleatório não observável. Dessa forma, toda informação sobre o parâmetro desconhecido θ é incerto, e tal incerteza é quantificada em termos de probabilidade.

Assim como a estatística clássica, a estatística Bayesiana também considera as informações contidas nos dados (na amostra) através da função de verossimilhança. Porém o maior ganho do método está em considerar, ou melhor, não ignorar qualquer informação inicialmente conhecida sobre o parâmetro através da probabilidade *a priori*. Dessa forma, podemos defini-la como uma distribuição de probabilidade para θ, normalmente subjetiva, que exprime o grau

de credibilidade que o indivíduo (ou pesquisador que faz à análise) atribui ao particular valor considerado a θ (JAYNES, 1996). Desse modo, tudo isso é traduzido na linguagem matemática através do Teorema de Bayes que possibilita unir todos esses conceitos em uma expressão simples.

2.2.1 Teorema de Bayes

O Teorema de Bayes é uma probabilidade condicional indubitável tanto para a estatística clássica quanto a Bayesiana, uma vez admissível as leis tradicionais do cálculo de probabilidade e os axiomas de Kolmogorov. O centro da discussão sobre o tema se encontra na interpretação conceitual e na sua aplicação a problemas de inferência estatística. Dessa forma, o completo entendimento sobre o Teorema de Bayes é fundamental para a inferência Bayesiana.

Como critério e rigor matemático, definiremos um espaço de probabilidade pelo terno $(\Omega, \mathscr{A}, P)$, em que: Ω é um espaço não vazio (espaço amostra) com elementos $\omega \in \Omega$ e subconjuntos $A \subseteq \Omega$, designados como eventos do espaço amostral. $\mathscr{A}$ é a família (σ-álgebra) dos eventos dotados de probabilidade. P é a medida de probabilidade definida para os eventos de $A \subseteq \Omega$, $A \in \mathscr{A}$, em que $P(A)$ é a probabilidade do acontecimento de A.

Considere uma partição finita (ou infinita) de Ω (Figura 2.3)

$$A_1, A_2, \ldots, A_m;\ P(A_i) > 0;\ A_i \cap A_j = \emptyset\ \forall\ i \neq j,\ \text{portanto,} \bigcup_{i=1}^{m} A_i = \Omega.$$

Dado um outro evento B qualquer, com $P(B) > 0$, podemos decompor em eventos disjuntos onde sua união será o próprio evento B,

$$B = \bigcup_{i=1}^{m} (A_i \cap B).$$

Consequentemente, do axioma de aditividade de *Kolmogorov* da função P e a própria definição de probabilidade condicional, tem-se,

$$P(B) = \sum_{i=1}^{m} P(A_i \cap B) = \sum_{i=1}^{m} P(B \mid A_i) P(A_i).$$

Tomando qualquer evento i, tem-se em nota,

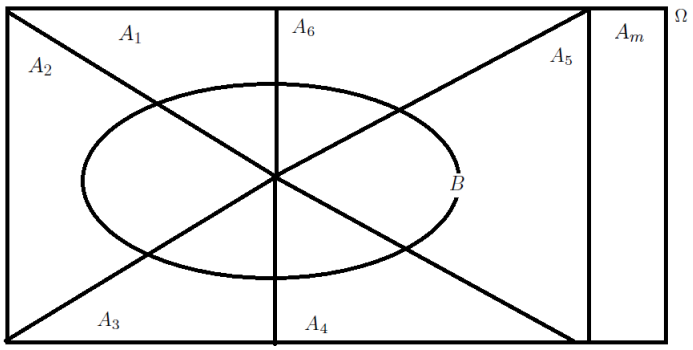
$$P(A_i \cap B) = P(B \mid A_i) P(A_i) = P(A_i \mid B) P(B),$$

e resolvendo em função de $P(A_i \mid B)$, define-se o Teorema de Bayes:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)} = \frac{P(B \mid A_i)P(A_i)}{\sum_{i=1}^{m} P(B \mid A_i)P(A_i)}.$$

É válido ressaltar que o Teorema de Bayes é um dos poucos resultados da estatística matemática que se propõe em caracterizar a aprendizagem com a experiência, isto é, a atualização da informação inicial acrescida de novos conhecimentos (acontecimentos) após a experiência ou a observação. Em especial a essa probabilidade $P(A_i \mid B)$ denominaremos, em futuras sessões, de distribuição *a posteriori*. No entanto, no momento, considere $P(A_i \mid B)$ apenas como uma probabilidade condicional de uma partição finita ou infinita de Ω.

Figura 2.3 – Figura ilustrando a partição de um espaço de probabilidade não vazio definida $(\Omega, \mathscr{A}, P)$.



Fonte: criação própria (2023)

2.2.2 Distribuição *a priori*

A distribuição *a priori* é uma caracterização distributiva da crença sobre um conjunto de parâmetros antes de visualizar os dados. Segundo a descrição dada por Box e Tiao (1992), a distribuição *a priori* resume a informação relativa ao parâmetro θ, até então "desconhecido" antes da realização do experimento, isto é, modela a incerteza relativa ao conjunto de parâmetros antes da observação dos dados.

Denotaremos a distribuição *a priori* como sendo $h(\theta)$. Faremos um paralelo ao Teorema de Bayes, porém levando em conta a função densidades de probabilidade. Suponha $Y = y$ um evento ocorrido e considere qualquer elemento de $\mathscr{F}$ como por exemplo $f(y \mid \theta)$. A distribuição *a priori* $h(\theta)$, no Teorema de Bayes se apresenta na expressão,

$$h(\theta \mid y) = \frac{f(y \mid \theta)\, h(\theta)}{\int_{\Theta} f(y \mid \theta)\, h(\theta)\, d\theta},\ \theta \in \Theta \tag{2.1}$$

em que $h(\theta \mid y)$ é a distribuição *a posteriori* de θ dado o evento $Y = y$ ocorrido. Dessa forma, a informação contida nos dados y mais a representação do investigador caracterizada por $h(\theta)$, é atualizada e representada por $h(\theta \mid y)$.

A distribuição *a priori* de forma geral é subjetiva envolvendo julgamentos e experiências individuais, por essa razão $h(\theta)$ é designada como graus de credibilidades para essa distribuição com base no conhecimento de θ até o momento. Definir probabilidades subjetivas não é tarefa fácil para o pesquisador, que por sua vez é tema de conflito para a estatística clássica ao questionar a imprecisão dessa abstração, a replicabilidade e a elucidação do método científico. Esse impasse é debatido e fundamentado (pela estatística Bayesiana) por métodos baseados no princípio da consistência, que possibilita definir distribuições *a priori* em caráter científico.

Existem muitos métodos para obter distribuições *a priori*. Os mais adotados são as **famílias de distribuições conjugadas**, que por algumas conveniências analíticas se define distribuições apropriadas ao fenômeno elucidado (PAULINO; TURKMAN; MURTEIRA, 2003); **distribuições *a priori* subjetivas**, que quantifica informações de natureza subjetivas transformando em uma distribuição de probabilidade. Podemos listar um gama de pesquisas na literatura que tratam sobre o assunto como French (1985), Lindley (1983), Lindley (1985), Genest e Schervish (1985), West e Crosse (1992), Phillips e Wisbey (1993), Gelfand, Mallick e Dey (1995), Kadane e Wolfson (1998), O'Hagan et al. (2006), Chaloner (1996), Garthwaite, Kadane e O'Hagan (2005), Oakley e O'Hagan (2007) e Johnson et al. (2010); e quando não se têm quaisquer informações ou essas são praticamente escassas, utiliza-se **distribuições não informativas** que é um estado probabilístico da representação conceitual lógico da incerteza na forma de probabilidade. Os métodos não informativo são vários entre eles o Método de Bayes-Laplace, Método de Jeffreys e Método de Box-Tiao, com devida atenção às distribuições *a priori* impróprias que por sua vez podem levar a distribuições *a posteriori* impróprias impossibilitando fazer qualquer inferência sobre o conjunto paramétrico. Entretanto, todas elas devem satisfazer o princípio de coerência ou de consistência permitindo o uso do cálculo de probabilidade e seus axiomas já estabelecidos.

2.2.3 Função de verossimilhança

A teoria de verossimilhança desempenha papel de destaque na inferência clássica e (como é notório no teorema de Bayes) na estatística Bayesiana. Antes de exemplificá-la podemos defini-la como uma função dos parâmetros de um modelo estatístico que permite inferir

sobre os valores dos parâmetros a partir de um conjunto de observações. Ou seja, digamos que se tenha conhecimento ou suposição da distribuição de probabilidade de uma variável aleatória qualquer, porém, gostaríamos de quantificar os valores dos parâmetros que governam o fenômeno estudado. Através do método de máxima verossimilhança é possível inferir o valor mais plausível sobre esses parâmetros considerando apenas a amostra observada. Segue a definição formal.

Definição: Uma sequência de $Y_1, Y_2, \ldots, Y_n$ de n variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com função densidade de probabilidade (*f.d.p.*) ($f(y \mid \theta)$) ou, no caso discreto, função de massa de probabilidade (*f.m.p.*) ($P(y \mid \theta)$) é dita ser uma amostra aleatória de tamanho n da distribuição de Y. Nesse caso, temos,

$$f(y_1, y_2, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} f(y_i \mid \theta) = f(y_1 \mid \theta) f(y_2 \mid \theta) \cdots f(y_n \mid \theta).$$

Dessa forma, concluímos que usamos a amostra $Y_1, Y_2, \ldots, Y_n$ para obter informação sobre o vetor de parâmetros θ. A função densidade conjunta dada pela definição é denominada função de verossimilhança de θ, correspondente à amostra observada $Y = (y_1, y_2, \ldots, y_n)'$ e será denotada por

$$L(\theta; y) = \prod_{i=1}^{n} f(y_i \mid \theta).$$

Qualquer função da amostra que não depende de parâmetros desconhecidos é denominada uma estatística. O estimador $\hat{\theta} \in \Theta$ é uma estatística pois é um estimador para θ em função da amostra que não depende do parâmetro. Assim, o valor mais provável para $\hat{\theta}$ dada a amostra observada será

$$\hat{\theta} = sup\, L(\theta; y).$$

A função chamada de verossimilhança $L(\theta; y) = f(y \mid \theta)$ é condicional aos dados observado. Ela deve ser interpretada como uma função de θ (vetor de parâmetros) baseada na informação dos dados (fixos) observados. Ademais, ela serve para medir o quanto os dados suportam uma hipótese sobre θ.

2.2.4 Distribuição *a posteriori*

Em análise Bayesiana, inferências são realizadas diretamente da distribuição *a posteriori*. A partir dela podem-se obter estimativas pontuais, obtidas para resumir as características de

tal distribuição, como também estimativas intervalares e inferências probabilísticas conhecida como intervalos de credibilidade, como mencionado anteriormente.

No Teorema de Bayes, a distribuição *a posteriori* é proporcional à verossimilhança e *a priori*,

$$h(\theta \mid y) \propto f(y \mid \theta)\, h(\theta).$$

A integração do denominador observado na Equação 2.1 é a distribuição marginal de y em relação à densidade conjunta e funciona como um normalizador para que seu resultado mantenha-se a condição de: i) $h(\theta \mid y) > 0$, ii) $\int h(\theta \mid y)\, \mathrm{d}\theta = 1$ (LARSON, 1982).

2.3 Distribuições Hierárquicas

Frequentemente, as observações têm algum tipo de hierarquia natural, de modo que elas podem ser agrupadas e modeladas pertencendo a grupos diferentes (*e.g.* os resultados de uma pesquisa podem ser agrupados a nível de país, região, estado, cidades, município ou mesmo bairro). Esses diferentes grupos, por sua vez, também podem ser modelados como membros de subgrupo comum, e assim sucessivamente. Dessa forma, essas estruturas hierárquicas das quantidades observáveis correspondem a uma decomposição do modelo amostral em dois ou mais níveis.

A ideia da modelagem hierárquica é usar os dados para modelar a força da dependência entre os grupos. Presume-se que os grupos sejam uma amostra da distribuição populacional subjacente e a variação dessa distribuição populacional (estimada a partir dos dados) determina quanto os parâmetros da distribuição amostral são reduzidos em relação à média comum.

Normalmente as distribuições *a priori* hierárquicas são especificadas em dois ou três níveis, embora não haja limitação quanto ao número dos níveis (GELMAN, 2006). No entanto quanto mais níveis a hierarquia tiver maior será o grau de complexidade resultante do modelo.

Para que o leitor compreenda melhor em detalhes vamos a um exemplo. Considere um modelo hierárquico de dois níveis sendo J grupos diferentes e $n_1, \ldots, n_J$ observações de cada um dos grupos, podendo ser escrito como

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \text{ para todo } i = 1, \ldots, n_j$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \text{ para todo } j = 1, \ldots, J$$

para cada $j = 1,\ldots,J$ grupo. É válido ressaltar que ϕ é denominado de hiperparâmetro do modelo, que está no segundo nível da hierarquia nesse exemplo.

Assumimos que as observações $Y_{1j},\ldots,Y_{n_{ij}}$ dentro da cada grupo são variáveis aleatórias independentes e identicamente distribuídas (*i.i.d.*), de modo que a distribuição amostral conjunta pode ser escrita como um produto das distribuições amostrais das observações:

$$p(y_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

Os parâmetros dos grupos $(\theta_1,\ldots,\theta_J)'$ são então modelados como uma amostra *i.i.d.* da distribuição populacional comum $p(\theta_j \mid \phi)$ de modo que sua distribuição conjunta também pode ser fatorada como

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

Assim, a distribuição *a priori* de um parâmetro θ dependerá dos valores dos hiperparâmetros. Portanto a especificação completa do modelo depende de como lidamos com os hiperparâmetros. Existem ao menos três formas de lidar com eles:

1. Fixa-los a alguns valores constantes;
2. Usar estimativas pontuais a partir dos dados;
3. Definir uma distribuição de probabilidade sobre eles.

Devido à dificuldade de interpretação dos hiperparâmetros em níveis mais altos, na maioria dos casos a terceira opção é a mais utilizada. Deste modo é comum especificar distribuições *a priori* subjetivas para esses hiperparâmetros nestes níveis (BERNARDO; SMITH, 1994). Abaixo detalharei apenas o terceiro método que é o necessário para a perfeita compreensão desta Tese.

2.3.1 Modelo Hierárquico totalmente Bayesiano

Para especificar o modelo totalmente Bayesiano, definimos uma distribuição prévia também para os hiperparâmetros, para que o modelo completo se torne:

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \text{ para todo } i = 1,\ldots,n_j$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \text{ para todo } j = 1,\ldots,J$$

$$\phi \sim p(\phi)$$

Já fizemos explicitamente as seguintes suposições de independência condicional:

$$Y_{11}, \ldots, Y_{n_1 1}, \ldots, Y_{1J}, \ldots, Y_{n_J J} \perp\!\!\!\perp | \ \theta$$

$$\theta_1, \ldots, \theta_J \perp\!\!\!\perp | \ \phi,$$

mas a crucial suposição implícita da independência condicional ($\perp\!\!\!\perp$[1]) do modelo hierárquico é que os dados dependem dos hiperparâmetros somente através dos parâmetros no nível da população:

$$Y \perp\!\!\!\perp \phi \mid \theta.$$

Isso significa que a distribuição amostral das observações, dados os parâmetros populacionais se resume a

$$p(y \mid \theta, \phi) = p(y \mid \theta),$$

e assim a distribuição *a posteriori* completa sobre os parâmetros pode ser escrita usando a fórmula de Bayes:

$$p(\theta, \phi | y) \propto \ p(y|\theta, \phi) \ p(\theta, \phi) \tag{2.2}$$

$$p(\theta, \phi | y) \propto p(y|\theta) \ p(\theta|\phi) \ p(\phi) \tag{2.3}$$

$$p(\theta, \phi | y) \propto p(\phi) \prod_{j=1}^{J} p(y_j|\theta_j) \ p(\theta_j|\phi). \tag{2.4}$$

Como agora *a posteriori* completa não é mais fatorada, não podemos resolver *as posteriores* marginais dos parâmetros em nível de grupo $p(\theta_j \mid y)$ independentemente. Dessa forma, todo o modelo não pode ser resolvido analiticamente. Portanto, para solucionar esse problema os métodos computacionais são muito utilizados nesse tipo de modelo.

Na abordagem totalmente Bayesiana, a distribuição marginal *a posteriori* dos parâmetros em nível de grupo é obtida pela integração da distribuição *a posteriori* condicional dos parâmetros em nível de grupo sobre toda a distribuição posterior marginal dos hiperparâmetros (ou seja, considerando o valor esperado da distribuição *a posteriori* condicional dos parâmetros

[1] Símbolo matemático de independência condicional de variáveis aleatórias na teoria da probabilidade.

em nível de grupo sobre *a posteriori* marginal dos hiperparâmetros):

$$p(\theta|y) = \int p(\theta,\phi|y)\ \mathrm{d}\phi = \int p(\theta|\phi,y)\ p(\phi|y)\ \mathrm{d}\phi.$$

Isso significa que o modelo totalmente Bayesiano leva em consideração adequadamente a incerteza sobre os valores dos hiperparâmetros, calculando o seu valor pela média a partir da distribuição *a posteriori*.

2.4 Métodos computacionais

Um dos grandes entraves para a generalização do método Bayesiano no século XX é a resolução de complexas integrais necessárias para se fazer a inferência Bayesiana. Desde os anos 80 investigações têm se concentrado em técnicas eficientes para superar essa dificuldade. Métodos de aproximações da distribuição *a posteriori* por uma distribuição Normal multivariada foram abordado por Laplace, até evoluir para métodos mais sofisticados como quadraturas numéricas iterativas sugerido por Naylor e Smith (1982), métodos de Monte Carlo e Métodos de Monte Carlo via Cadeias de Markov (MCMC).

Atualmente o principai algoritmo numérico utilizado, quando os cálculos analíticos são muito complexos ou impossíveis de serem obtidos, é o algoritmo de Metropolis-Hastings, criado e investigado por Metropolis et al. (1953), Hastings (1970), Geman e Geman (1984). Sua relação com a estatística Bayesiana para solucionar problemas de integrais complexas e consequentemente obter distribuições *a posteriori* marginais, foi mencionado pela primeira vez por Gelfand e Smith (1990) e se tornou uma técnica extremamente poderosa que transformou as investigações no domínio do universo Bayesiano.

Destaco que existem os modelos conjugados para os quais é possível resolver a função de verossimilhança marginal e, portanto, também as distribuições *a posteriori* e suas predições de forma fechada. No entanto, em cenários mais realistas nos quais são necessários modelos mais complexos, as verossimilhanças marginais são geralmente intratáveis e por isso *a posteriori* não pode ser resolvida analiticamente.

Isso significa que geralmente temos que aproximar a distribuição posterior $p(\theta \mid y)$ de alguma forma, e usar essa aproximação para calcular as quantidades de interesse, como média posteriori ou intervalos de credibilidade.

Dessa forma os próximos tópicos descrevem uma breve revisão introdutória das principais técnicas de aproximação computacionais desenvolvidas para decifrar o paradigma Bayesiano. No entanto, dar-se-á mais ênfase ao Método Monte Carlo Hamiltoniano devido a base para o entendimento completo dessa tese.

2.4.1 Simulação de Monte Carlo

Assim como toda simulação, o primeiro passo é gerar amostras aleatórias do espaço paramétrico. A grande dúvida é: se não conhecemos a distribuição a posteriori como é possível gerar amostras a partir de uma distribuição desconhecida?

Técnicas computacionais de integração por simulações são conhecidas como Integração de Monte Carlo ou Método de Monte Carlo. Elas se baseiam no teorema clássico de probabilidade muito conhecido como Lei Forte dos Grandes Números.

Seja $Y_1, Y_2, \ldots, Y_n$ variáveis *i.i.d.* com um valor esperado $\mu = E[Y] < \infty$, ou seja um valor finito, teremos

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \mu$$

convergindo quase certo quando $n \to \infty$.

A convergência quase certa significa que a sequência converge com probabilidade igual a um. Outra maneira de indicar o resultado é

$$P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} Y_i = \mu\right) = 1.$$

2.4.2 Método de Monte Carlo via Cadeia de Markov (MCMC)

O método Monte Carlo via Cadeias de Markov (MCMC) compreende-se como uma classe de algoritmos baseados na amostragem iterativa de uma cadeia de Markov cuja distribuição estacionária é a distribuição alvo (ROBERT; CASELLA, 2004), no caso da computação Bayesiana é a distribuição *a posteriori* de interesse.

Já uma Cadeia de Markov é um processo estocástico no qual o próximo estado da cadeia depende somente do estado atual e dos dados, não do histórico sobre a cadeia (PAULINO; TURKMAN; MURTEIRA, 2003). Uma cadeia de Markov de tempo discreto é uma sequência de variáveis aleatórias $Y_1, Y_2, \ldots, Y_n$ que possui a seguinte propriedade

$$P(Y_{i+1} = y_{i+1} \,|\, Y_i = y_i, \ldots, Y_0 = y_0) = P(Y_{i+1} = y_{i+1} \,|\, Y_i = y_i)$$

para todo $i = 1, 2, \ldots, n$. Isso significa que a qualquer momento o estado futuro Y_{i+1} depende apenas do estado atual Y_i da cadeia, e não do resto do histórico como já foi mencionado.

Em consequência a essa dependência gera-se uma forte autocorrelação entre as observações subsequentes da cadeia $\theta_1, \theta_2, \ldots, \theta_{\mathbf{S}}$ amostrados sobre o espaço paramétrico Θ. Isso significa que, muito provável, o próximo valor θ_{i+1} esteja próximo e correlacionado ao valor atual θ_i da cadeia. Para diminuir essa alta correlação existente entre os valores amostrais deve-se descartar alguns elementos entre uma iteração para outra, esse valor é denominado por janela (*lag*).

As primeiras iterações da amostragem MCMC geralmente são descartadas porque os valores da cadeia antes de convergir para a distribuição estacionária não são representativos da distribuição posterior. Essas iterações descartadas são chamadas de período de queima ou de aquecimento (*burn-in*) e a quantidade de pontos descartados é uma questão de escolha. Existem alguns métodos para essa decisão porém a linguagem *Stan* (que utiliza método Hamiltoniano Monte Carlo com o amostrador NUTS - *No-U-Turn Sample*, detalhado no capítulo dedicado posteriormente) define esse período de aquecimento automaticamente para que o analista não se preocupe com isso.

A inferência Bayesiana depende da qualidade da convergência e da amostragem da simulação, portanto, existem métodos de diagnóstico do modelo para medir a qualidade dessa convergência. Mais uma vantagem da linguagem *Stan* é que ela possui uma série de métricas bastante avançadas e eficientes que servem de diagnóstico do modelo. Uma delas é executar uma série de cadeias a partir dos diferentes valores iniciais em paralelo. Se todas convergirem para uma distribuição semelhante, é bem provável que essa seja a distribuição estacionária. *Stan* executa quatro cadeias paralelas como padrão.

Dentre todos os amostradores MCMC os mais populares são: o amostrador Metropolis-Hastings e o amostrador Gibbs (que também pode ser visto como um caso especial do amostrador Metropolis-Hasting).

2.4.3 Método Metropolis e Metropolis-Hastings

O algoritmo Metropolis-Hastings é um termo geral para uma família de métodos de simulação em cadeia de Markov com a finalidade de extrair amostras das arbitrárias distribuições de probabilidade *a posteriori* Bayesianas e explorar o espaço paramétrico desconhecido.

O algoritmo Metropolis é uma caminhada aleatória (*random walk*) pelo espaço paramétrico que usa uma regra de aceitação/rejeição a cada iteração para convergir a específica distribuição alvo. Gera-se uma amostra da distribuição conjunta *a posteriori* $\pi(\theta \mid t)$, a partir de um valor aleatório do parâmetro candidato ou proponente $\theta^{(c)}$ passando por um crivo, com uma dada probabilidade de aceitação, a cada transição da cadeia (HASTINGS, 1970; METROPOLIS et al., 1953).

- **Algoritmo Metropolis-Hastings resumido:**

1. Inicialize o contador de iterações $t = 0$ e especifique os valores iniciais $\theta^{(0)} = \theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_d^{(0)}$ para os parâmetros;

2. Gere um valor $\theta^{(c)}$ da distribuição proposta $q(\cdot \mid \theta_1)$, também conhecida como *função de importância*, distribuição instrumental ou distribuição proponente (por diferentes autores);

3. Calcule a probabilidade de aceitação $\alpha(\theta_1, \theta^{(c)})$, em que:

$$\alpha(\theta_1, \theta^{(c)}) = min\{\frac{\pi(\theta^{(c)} \mid \theta_2, \ldots, \theta_d) q(\theta_1 \mid \theta^{(c)})}{\pi(\theta_1 \mid \theta_2, \ldots, \theta_d) q(\theta^{(c)} \mid \theta_1)}\ ,\ 1\};$$

   interprete como o valor mínimo entre a *razão de densidade* $r = \frac{\pi(\theta^{(c)}|\theta_2,\ldots,\theta_d) q(\theta_1|\theta^{(c)})}{\pi(\theta_1|\theta_2,\ldots,\theta_d) q(\theta^{(c)}|\theta_1)}$ e a probabilidade máxima 1,

4. Gere um valor u, a partir de uma distribuição uniforme $U(0,1)$;

5. Se $u < \alpha$, então aceite o valor candidato e faça $\theta_1^{(t+1)} = \theta^{(c)}$. Caso contrário, rejeite e faça $\theta_1^{(t+1)} = \theta^{(t)}$;

6. Incremente o contador de t para $t+1$ e volte ao passo (2) até atingir a convergência.

É válido mencionar que a incidência de convergência da cadeia não significa um perfeito desempenho do algoritmo e consequentemente um representativo conjunto de estado da distribuição *a posteriori*, mesmo em um tempo relativamente rápido computacional. Uma função de importância $q(\cdot \mid \cdot)$ adequada deve produzir valores varrendo o espaço paramétrico em um relativo número de iteração com uma porcentagem equilibrada de valores proponentes aceitos e rejeitados (GELMAN et al., 2013).

A taxa de aceitabilidade e rejeitabilidade é uma ótima ferramenta de diagnóstico de convergência e funcionamento do algoritmo que está relacionado com a dispersão da distribuição proposta para gerar a simulação. Se q for muito disperso em relação a π, valores produzidos são muitos rejeitados, obtendo uma amostra representativa possivelmente após muitas iterações. Em contrapartida, com uma dispersão pequena de q apenas um pequeno subespaço amostral poderá ser varrido, em um número grande de iterações, com passos curtos proporcionando uma alta taxa de aceitação e uma equivocada convergência rápida (PAULINO; TURKMAN; MURTEIRA, 2003). Caso o leitor deseje uma visualização gráfica e uma explicação mais visual e didática sobre o assunto, uma série de vídeos no YouTube (canal: Carlos Zarzar https://www.youtube.com/@carloszarzar/videos) foram disponibilizados para maiores interessados.

Pelo fato de ser um método de Monte Carlo via Cadeia de Markov o qual gera amostras aleatórias dependentes por uma distribuição proposta, espera-se que haja uma alta autocorrelação as observações subsequentes da cadeia, podendo diminuir esse efeito considerando um espaçamento entre as iterações armazenadas denominado de *lag* ou *thinning interval* como mencionado anteriormente.

2.4.4 Método de Gibbs

Um algoritmo pertencente a família MCMC que foi considerado útil em muitos problemas multidimensionais é o amostrador de Gibbs. É um método de amostragem rápida que pode ser usado em situações em que as distribuições condicionais são conhecidas. Também chamado de amostragem por distribuições condicionais completas (PAULINO; TURKMAN; MURTEIRA, 2003), é definida em termos de subvetores θ.

Suponha que o vetor de parâmetro θ tenha sido dividido em d componentes ou subvectores, $\theta = (\theta_1, \ldots, \theta_d)'$. Cada iteração do amostrador de Gibbs circula pelos subvetores de θ, desenhando cada subconjunto condicional sobre os valores dos demais. Portanto, existem d etapas na iteração t. Em cada iteração t, uma ordem dos d subvectores de θ é escolhida e, por sua vez, cada $\theta_j^{(t)}$ é amostrado da distribuição condicional, considerando todos os outros componentes de θ:

$$\pi(\theta_j \mid \theta_{-j}^{(t-1)}, y),$$

conhecida como função de suporte, em que $\theta_{-j}^{(t-1)}$ representa todos os componentes de θ, exceto θ_j, com seus valores atualizados:

$$\theta_{-j}^{(t-1)} = (\theta_1^{(t)}, \ldots, \theta_{j-1}^{(t)}, \theta_{j+1}^{(t-1)}, \ldots, \theta_d^{(t-1)}).$$

Assim, cada subvetor θ_j é atualizado condicionalmente aos valores mais recentes dos outros componentes de θ, que são os valores da iteração t para os componentes já atualizados e os valores da iteração $t-1$ para os demais (GELMAN et al., 2013).

As transições entre os estados da cadeia serão feitas a partir de distribuições condicionais completas $\pi(\theta_j \mid \theta_{-j}^{(t-1)}, y)$, em que $\theta_{-j}^{(t-1)}$ é um vetor como já mencionado, o qual seus elementos θ_{-j} podem ser unidimensionais ou multidimensionais (ROBERT; CASELLA, 2013).

Se as distribuições condicionais forem completamente conhecidas, então o algoritmo de Gibbs pode ser dado a partir da obtenção de $\theta^{(t)}$, por meio da especificação de $\theta^{(t-1)}$, através da geração sucessiva de valores

$$\begin{aligned}
\theta_1^t &\approx \pi(\theta_1 \mid \theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)}) \\
\theta_2^t &\approx \pi(\theta_2 \mid \theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)}) \\
&\vdots \\
\theta_d^t &\approx \pi(\theta_d \mid \theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{d-1}^{(t)}).
\end{aligned}$$

Cada iteração se completa após d passos ao longo do vetor θ. No final das iterações, ocorre a convergência, os valores resultantes irão compor a amostra de $\pi(\theta)$. Assim, o amostrador de Gibbs é considerado por muitos um caso especial do algoritmo de Metropolis-Hastings, onde os elementos de θ são atualizados com base na distribuição condicional completa e a probabilidade de aceitação é igual a 1. Portanto, não existe um mecanismo de aceitação pois a cadeia sempre irá se mover para um novo valor (sempre ocorrerá aceitação).

- **Algoritmo Gibbs resumido:**

1. Fixar um conjunto de valores iniciais $\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_d^{(0)}$;
2. Obter um novo valor $\theta^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_d^{(t)})$ a partir de um valor $\theta^{(t-1)}$ através da geração sucessiva de valores dada anteriormente;

3. Mudar o contador de t para $t+1$ e voltar a etapa (2) até obter a convergência.

O fato de que o algoritmo de Gibbs, necessariamente, simula cada iteração para um componente do vetor θ gerando uma função suporte $\pi(\theta_j \mid \theta_{-j}^{t-1}, y)$ relativamente mais lenta, consequentemente, há uma retardação na convergência comparado com outros algoritmos (PAULINO; TURKMAN; MURTEIRA, 2003). Isso ocorre principalmente para modelos com grandes dimensões como os modelos hierárquicos.

Na prática, tem-se demonstrado que em alguns modelos o algoritmo Metropolis-Hastings são mais eficientes do que o amostrador de Gibbs, para situação de complexas distribuições condicionais completas (PAULINO; TURKMAN; MURTEIRA, 2003). Entretanto, existe o risco factual da distribuição *a posteriori* proveniente do algoritmo de Metropolis-Hastings não corresponder a distribuição alvo, se colocando em posição desvantajosa com relação ao Gibbs.

Para os leitores interessados em uma visualização gráfica e uma explicação mais visual e didática sobre o algoritmo Gibbs, uma série de vídeos no YouTube (canal: Carlos Zarzar https://www.youtube.com/@carloszarzar/videos) foram disponibilizados para maiores compreensões sobre o assunto.

2.4.5 Método Hamiltoniano Monte Carlo

O Hamiltoniano Monte Carlo (HMC) é um método de Monte Carlo via cadeia de Markov (MCMC) que usa derivadas de função densidades, amostradas para gerar transições eficientes que investiguem todo o espaço paramétrico desconhecido definindo a densidade *a posteriori* da melhor forma possível. Ele usa simulação aproximada da dinâmica Hamiltoniana baseada na integração numérica corrigida executando uma etapa de aceitação do algoritmo de Metropolis.

O HMC seguiu um longo e sinuoso caminho até a computação estatística moderna. O método foi originalmente desenvolvido no final da década de 80 inicialmente chamado de Híbrido Monte Carlo para lidar com os cálculos na Cromodinâmica Quântica Lattice (DUANE et al., 1987). Dentro de alguns anos, Radford Neal enxergaria o potencial do método para problemas em estatística aplicada no seu pioneiro trabalho em Redes Neurais Bayesianas (NEAL, 1995). Na década seguinte, o método começou a aparecer em livros didáticos tornando-se mais acessível a outros pesquisadores até MacKay e Kay (2003) e posteriormente Bishop (2006) usar o termo Método Hamiltoniano Monte Carlo.

A influente revisão de Neal (NEAL et al., 2011) foi a que realmente introduziu a abordagem na computação estatística. Com o surgimento de implementações de software de alto

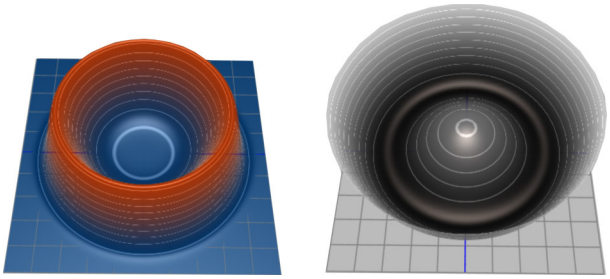
desempenho, como Stan (STAN, 2017), o método agora se tornou uma ferramenta difundida em muitas aplicações científicas, médicas e industriais.

- **Analogia física/mecânica por trás do método HMC**

O grande diferencial do algoritmo HMC é não precisar diretamente de uma amostragem totalmente aleatória no espaço paramétrico, mas de uma aleatoriedade direcionada, guiada, ou definida, com um princípio de condução. Um amostragem *random walk* é interessante, pois pressupõe que o espaço seja maximamente varrido da forma mais imparcial possível, no entanto, o custo computacional para essa equanimidade, principalmente para modelos mais complexos, pode ser muito alto. Um método aleatório que consiga amostrar mais o pico (ou os picos no caso de distribuições multimodais, porém sem deixar de cobrir todo o espaço) da função densidade de probabilidade alvo do que andar em todo o domínio não representativo, seria muito mais interessante e de menor custo computacional.

Em vez de fazer propostas aleatórias, suponha que o algoritmo execute uma simulação física do espaço paramétrico alvo, considerando o vetor de parâmetros como uma partícula no espaço n-dimensional. A superfície neste espaço é irregular (com vales e montes) determinada pela função densidade de probabilidade (*a posteriori*) transformada (logaritmizada) e espelhada com sinal negativo (reflexão em relação ao um eixo imaginário) (Figura 2.4).

Figura 2.4 – Analogia física do espaço Hamiltoniano. A) Distribuição *a Posteriori* desconhecida; B) espaço paramétrico desconhecido transformado e espelhada simulando uma superfície gravitacional



Fonte: Site Alex Rogozhnikov (2023)

Suponha que essa superfície não tenha atrito e que o amostrador deslise nesse espaço em uma direção aleatória, sofrendo o efeito gravitacional, coletando informação ao longo de sua trajetória antes de outro lançamento dessa partícula simulada. Assim, no final de várias iterações pode-se inferir algo sobre a forma da superfície por onde a partícula deslizou. Considerando o mecanismo de aceitação muito utilizado por Metropolis-Hastings, as propostas estão dentro

da região de alta probabilidade da distribuição alvo, consequentemente, menos propostas são rejeitadas podendo se afastar do ponto de partida para que a cadeia explore eficientemente todo o formato do alvo em menos tempo. Efetivamente, ele flui através da distribuição a posteriori e mapeia toda a sua forma muito mais rapidamente quando comparado com demais algoritmos.

Para os leitores interessados em uma visualização gráfica e uma explicação mais visual e didática sobre o algoritmo, um vídeo no YouTube (canal: Carlos Zarzar) foi disponibilizado para maiores compreensões sobre o assunto.

- **Variável auxiliar momento**

O algoritmo introduz uma variável auxiliar análoga ao momento da física ρ, nos parâmetros da distribuição alvo θ, com densidade conjunta

$$p(\rho,\theta) = p(\rho|\theta)p(\theta).$$

Na maioria das aplicações do HMC, incluindo Stan, a densidade da variável auxiliar (momento) é um normal multivariado que não depende dos parâmetros θ,

$$\rho \sim \mathsf{MultiNormal}(0,M).$$

M é a métrica euclidiana. Ela pode ser vista como uma transformação do espaço paramétrico que torna a amostragem mais eficiente, veja Betancourt (2017) para mais detalhes.

- **Função Hamiltoniana**

Uma vez definida a densidade condicional do momento, a densidade conjunta $p(\rho,\theta)$ a seguir define a função Hamiltoniana:

$$H(\rho,\theta) = -\log p(\rho,\theta)$$

$$H(\rho,\theta) = -\log p(\rho|\theta) - \log p(\theta)$$

$$H(\rho,\theta) = T(\rho|\theta) + V(\theta),$$

em que o termo $T(\rho|\theta) = -\log p(\rho|\theta)$ é chamado de energia cinética, e o termo $V(\theta) = -\log p(\theta)$ é chamado de energia potencial (a energia potencial é especificada pelo *Stan* através da definição de uma densidade de log).

- **Gerando transições para as equações Hamiltoniana**

Essa função hamiltoniana gera uma transição amostrando primeiramente o momento auxiliar. A partir do valor atual dos parâmetros θ, uma transição para um novo estado é gerada em dois passos antes de ser submetida a uma etapa de aceitação do algoritmo Metropolis. Primeiro, um valor para o momento é desenhado independentemente dos valores atuais do parâmetro,

$$\rho \sim \mathsf{MultiNormal}(0, M).$$

Em seguida, o sistema conjunto (θ, ρ) composto pelos valores atuais dos parâmetros θ e o novo impulso ρ é determinado através das equações Hamiltonianas,

$$\frac{d\theta}{dt} = +\frac{\partial H}{\partial \rho} = +\frac{\partial T}{\partial \rho}$$

$$\frac{d\rho}{dt} = -\frac{\partial H}{\partial \theta} = -\frac{\partial T}{\partial \theta} - \frac{\partial V}{\partial \theta}.$$

Com a densidade do momento é independente da densidade alvo, ou seja $p(\rho|\theta) = p(\rho)$, o primeiro termo na derivada do momento $\partial T/\partial \theta$ é zero, produzindo ao par (θ, ρ) as seguintes equações diferenciais em relação ao tempo,

$$\frac{d\theta}{dt} = +\frac{\partial T}{\partial \rho}$$

$$\frac{d\rho}{dt} = -\frac{\partial V}{\partial \theta}.$$

- **Integrador *Leapfrog***

Para resolver a equação diferencial Hamiltoniana dos dois estados, a linguagem *Stan* e outras implementações HMC usam o integrador *Leapfrog*. Ele é um algoritmo de integração numérica especificamente adaptado para fornecer resultados estáveis para os sistemas de equações Hamiltonianos.

Como a maioria dos integradores numéricos, o *Leapfrog* executa etapas discretas de um pequeno intervalo de tempo ε. O algoritmo de salto começa desenhando um novo valor de momento independentemente dos parâmetros θ ou o valor anterior valor do momento.

$$\rho \sim \mathsf{MultiNormal}(0, M).$$

Em seguida, alterna as atualizações de meio passo do momento e as atualizações de passo completos da posição.

$$\rho \leftarrow \rho - \frac{\varepsilon}{2}\frac{\partial V}{\partial \theta}$$

$$\theta \leftarrow \theta + \varepsilon M^{-1}\rho$$

$$\rho \leftarrow \rho - \frac{\varepsilon}{2}\frac{\partial V}{\partial \theta}.$$

Aplicando L passos no integrador *Leapfrog*, a total simulação terá tamanho $L \times \varepsilon$. O estado resultante no final da simulação serão indicadas (ρ^*, θ^*). Leimkuhler e Reich (2004) fornecem uma análise detalhada da integração numérica para sistemas Hamiltonianos, incluindo uma derivação do erro vinculado ao integrador de saltos.

- **Etapa de aceitação do Metropolis**

Se o integrador *Leapfrog* fosse perfeito numericamente, não seria necessário fazer mais randomização por transição do que gerar um vetor de momento aleatório. Em vez disso, o que é feito na prática para contabilizar erros numéricos durante a integração é aplicar uma etapa de aceitação do Metropolis, onde a probabilidade de manter a proposta (ρ^*, θ^*) gerado pela transição de (ρ, θ) é,

$$\min(1,\ \exp(H(\rho,\theta) - H(\rho^*,\theta^*))).$$

Se a proposta não for aceita, o valor do parâmetro anterior será retornado e usado para inicializar a próxima iteração.

- **Parâmetros do algoritmo HMC**

O algoritmo Hamiltoniano de Monte Carlo possui três parâmetros que devem ser definidos,

- discretização tempo ε;
- métrica M;
- número de passos dados L.

Na prática, a eficiência da amostragem, tanto em termos de velocidade da iteração quanto de iterações por amostra efetiva, é altamente sensível a esses três parâmetros de ajuste (NEAL et al., 2011; HOFFMAN; GELMAN, 2014).

Se ε for muito grande, o integrador da *Leapfrog* será impreciso e muitas propostas serão rejeitadas. Se ε for muito pequeno, uma infinidade de pequenos passos serão dados pelo integrador de salto, tomando longo tempo de simulação e recursos computacionais. Assim, o objetivo é equilibrar a taxa de aceitação entre esses extremos.

Se L for muito pequeno, a trajetória traçada em cada iteração será muito curta e a amostragem passará para uma caminhada aleatória. Se L for muito grande, o algoritmo terá muito trabalho em cada iteração consumindo novamente mais recursos computacionais que necessário.

Se a métrica M^{-1} é uma estimativa pobre da covariância da distribuição a posteriori, o tamanho do passo ε deve ser mantido pequeno para manter a precisão aritmética. Isso levaria a um grande L para compensar.

- **Tempo de integração**

O tempo de integração total, como já foi mencionado, é $L \times \varepsilon$ e está em função do número de etapas. Algumas interfaces para *Stan* definem um tempo t aproximado de integração e o intervalo de discretização (tamanho da etapa) ε. Nesses casos, o número de etapas será arredondado para baixo, conforme

$$L = \left\lfloor \frac{t}{\varepsilon} \right\rfloor .$$

2.4.6 HMC estático

Os princípios da dinâmica hamiltoniana se relacionam diretamente ao MCMC, fornecendo uma maneira de gerar transições eficientes. Algoritmos MCMC que utilizam dinâmica hamiltoniana são geralmente chamados de HMC. Dentre eles temos o HMC estático, um dos primeiros desenvolvidos (DUANE et al., 1987), sendo mais simples de compreender e contém a maioria das propriedades relevantes para entender o algoritmo HMC NUTS.

Como descrito anteriormente na analogia física/mecânica, o algoritmo HMC simula um objeto que desliza sobre uma superfície sem atrito (o qual, energia potencial e cinética permanecem constante no sistema) a partir de uma posição atual de momentos aleatórios por um período de tempo finito até uma posição final percorrendo uma trajetória sobre um espaço paramétrico desconhecido. No entanto, três fatores são relevantes a serem discutidos. O primeiro é como serão realizadas as simulações do movimento em *log-posteriori* arbitrários, ou seja, como serão gerados os caminhos aleatórios. Normalmente, modelos simples como uma parábola são utilizados porque possuem soluções analíticas para as equações diferenciais subjacentes e é possível

simular caminhos exatos e contínuos durante a trajetória. Todavia, lembramos que na prática para a maioria dos modelos, os caminhos contínuos são aproximados usando um método numérico conhecido como integrador de salto (*Leapfrog Integrator*). A trajetória dependerá do tamanho de passos (ε) e do número de passos (L). O vetor posição na etapa L é a amostra proposta para essa transição, enquanto as etapas intermediárias são descartadas. Quaisquer eventuais erros de aproximação fazem com que a partícula se desvie da trajetória e, portanto, a energia Hamiltoniana H (que é a energia potencial mais a cinética) não é constante ao longo do tempo, tornando-se uma ótima ferramenta de avaliação da qualidade da convergência das transições do algoritmo.

Um grande desafio para o usuário é determinar o comprimento ideal da trajetória (determinar os valores ε e L ideais). Se o comprimento da trajetória for muito curto, propostas distantes são impossíveis, levando a um passeio aleatório ineficiente. Se for muito longa, a trajetória irá refazer seus passos em uma trajetória circular ou elíptica, o que é um desperdício computacional. Assim, a eficiência depende do comprimento da trajetória, mas o comprimento ideal é difícil de determinar e uma etapa de ajuste é crucial e necessária para HMC estático (BETANCOURT, 2016). O mesmo comprimento pode ser alcançado realizando menos etapas de tamanho maior, ou mais etapas de tamanho menor. Como cada etapa é computacionalmente cara, quanto menos etapas, mais rápida será a transição. Porém, há uma desvantagem para tamanhos de etapa grandes: eles levam a uma maior variação em H e, em alguns casos, o erro de aproximação se acumula de modo que a energia total (H) vai para o infinito, conhecida como uma transição divergente. Dessa forma, a etapa de aceitação do algoritmo *Metropolis* é responsável pelo filtro dessas transições de energia, fazendo com que todos os estados propostos com a energia total menor do que o passo decorrente sejam aceitas com probabilidade $min(1, \alpha)$, em que α é o exponencial da energia perdida (probabilidade < 1 são aceitas). Aumentar o tamanho do passo reduz o tempo de execução, mas aumenta o erro de aproximação, levando a mais estados rejeitados e transições divergentes, degradando a eficiência do algoritmo. Otimizar o tamanho do passo é, portanto, outro passo crucial no HMC (BETANCOURT; BYRNE; GIROLAMI, 2014).

Uma vez determinado um tamanho de etapa (ε) e número de etapas L, a última etapa é especificar uma função de energia cinética. Em HMC, é normalmente a densidade logarítmica de um vetor aleatório normal multivariado onde a matriz de covariância é conhecida como matriz de massa. Anteriormente, assumíamos que a energia cinética era a soma dos momentos

quadrados, correspondendo a uma matriz de massa de identidade. O efeito da matriz de massa é transformar globalmente a distribuição a posteriori para ter uma geometria mais simples para amostragem. As variâncias (diagonal principal da matriz de massa) estendem-se para a posteriori para que todos os parâmetros tenham a mesma escala, enquanto as covariâncias transformam os parâmetros para que sejam aproximadamente independentes. Quando bem sucedidos, os parâmetros transformados têm uma escala de 1 e nenhuma correlação, assemelhando-se às variáveis aleatórias normais padrão $i.i.d$.

A matriz de massa é análoga à covariância da função proposta por vezes usado em amostradores Metropolis-Hastings, que pode ter um impacto substancial sobre a amostragem. Dependendo do modelo, os algoritmos HMC podem ser eficientes com uma matriz de massa de identidade, mas exigirão mais etapas de salto por transição e mais tempo. Assim, para obter uma amostragem eficiente com HMC, a matriz de massa deve se aproximar da covariância da posterior, mas essa informação muitas vezes não é conhecida *a priori*.

Especificar um comprimento de trajetória ideal, tamanho do passo e matriz de massa é fundamental para que o HMC estático funcione de forma eficiente, levando a exigir ajuste prático especializado e conhecimento *a priori* do fenômeno (NEAL et al., 2011). Felizmente, o NUTS automatiza esse processo e fornece amostragem eficiente com o mínimo ou nenhum ajuste.

2.4.7 Amostrador NUTS

O amostrador No-U-Turn (NUTS) estende o HMC estático ao automatizar o ajuste: nem o tamanho da etapa nem o número de etapas precisam ser especificados pelo usuário. NUTS determina o número de etapas por meio de um algoritmo de construção de árvore sofisticado. Uma única trajetória NUTS é construída acumulando iterativamente etapas. Na primeira iteração, uma única etapa de salto é feita a partir do estado atual, de forma que a trajetória tenha um total de duas etapas. Em seguida, mais duas etapas são adicionadas (total de quatro), depois mais quatro (total de oito) e assim por diante, com cada iteração dobrando o comprimento da trajetória. Este procedimento de duplicação se repete até que a trajetória volte sobre si mesma e ocorra um retorno, ou a trajetória divirja (isto é, H vai para o infinito). O número de duplicações é conhecido como profundidade da árvore. O aspecto principal desse algoritmo de construção de árvore é que ele cria automaticamente trajetórias que não são nem muito curtas nem muito

longas. Na prática, isso significa que os comprimentos das trajetórias variam entre as transições: pode levar 8 etapas ou 128, dependendo dos vetores de posição e o momento.

O amostrador NUTS determina o tamanho do passo adaptando-o durante a fase de aquecimento (*burn-in*) para uma taxa de aceitação alvo (*adapt delta* na linguagem Stan). O tamanho do passo ajustado é então usado para todas as iterações de amostragem. Em contraste com o HMC estático, NUTS não usa uma etapa de aceitação Metropolis, portanto, uma estatística análoga é usada para adaptação. Betancourt, Byrne e Girolami (2014) descobriram que essa taxa de aceitação alvo deve estar geralmente entre $0,6$ e $0,9$, com valores maiores sendo mais robustos na prática. Assim, o NUTS reduz efetivamente o HMC estático a um único parâmetro de ajuste especificado pelo usuário: a taxa de aceitação de destino.

2.5 Seleção de modelos Bayesianos

Eleger um modelo (ou um conjunto de modelos), entre uma família de modelos propostos, ajustados em um específico conjunto de dados para caracterizar um fenômeno, não é tarefa fácil. Primeiro, porque existe um número finito muito grande de modelos a serem ajustados além de suas diferentes parametrizações, o que se torna uma análise muito exaustiva ao compará-los e depois julgá-los, elegendo o vencedor como "o melhor" modelo ajustado ao fenômeno, nesse processo de modelagem. Em segundo lugar, como o estatístico inglês George Box afirmou pela primeira vez em um artigo publicado no *Journal of the American Statistical Association* em 1976: "Essencialmente, todos os modelos estão errados, mas alguns são úteis". Isso significa que todos os modelos são uma simplificação da realidade, porém, mesmo conhecendo parcialmente a realidade eles podem ser bastante úteis para entender, prever e explicar a complexidade do universo e auxiliar as ciências que utilizam dessa ferramenta para pesquisas e investigações.

Existem diversos métodos de seleção de modelos. Referente aos modelos Bayesianos, todas as inferências são resumidas pela distribuição a posteriori. Portanto, uma forma muito usual de avaliar um modelo, é por meio da verificação preditiva das posteriores do modelo resultante (RUBIN, 1984), ou seja, através da precisão de suas previsões (acurácia preditiva). No entanto, outros métodos também são adotados como a verificação preditiva da distribuição à priori (ao avaliar possíveis replicações envolvendo novos valores dos parâmetros), ou verificações mistas para modelos hierárquicos por exemplo (GELMAN; MENG; STERN, 1996).

Quando se têm um conjunto de modelos candidatos para modelagem, eles podem ser comparados usando o fator de Bayes, ou utilizando algum procedimento de aproximação mais prático (HOETING et al., 1999), ou até mesmo métodos de expansão contínuo do modelo (DRAPER, 1999). No entanto, os métodos mais recomendados, até então na literatura, para evitar um superajuste do modelo sobre os dados são a validação cruzada e os critérios de informação (AKAIKE, 1973; STONE, 1977). São abordagens que estimam a precisão preditiva fora da amostra usando ajustes dentro da amostra.

A validação cruzada exata requer o reajuste do modelo com diferentes conjuntos de treinamento. A validação cruzada aproximada deixando um de fora (conhecida no inglês como *leave-one-out cross-validation* LOO-CV) pode ser calculada facilmente usando a amostragem por importância (GELFAND; DEY; CHANG, 1992; GELFAND, 1996), porém, a estimativa resultante normalmente tem fortes ruídos, o que significa que a variância dos pesos por importância pode ser grande ou até infinito (PERUGGIA, 1997; EPIFANI et al., 2008). Felizmente, Vehtari, Gelman e Gabry (2017) propuseram o uso da amostragem por importância suavizada de Pareto (PSIS - *Pareto smoothed importance sampling*), uma nova abordagem que fornece uma estimativa mais precisa e confiável ajustando uma distribuição de Pareto à cauda superior da distribuição dos pesos de importância. PSIS permite calcular LOO usando pesos de importância que de outra forma seriam instáveis.

Dentre todos os critérios de informação, o critério de informação de Watanabe-Akaike, ou como ele mesmo nomeou o critério de informação amplamente aplicável (WATANABE; OPPER, 2010) (do inglês WAIC - *Widely Applicable Information Criterion*, ou também, *Watanabe Information Criterion*) é o mais recomendado para avaliar precisões preditivas de diferentes modelos, pois pode ser considerado um índice genuinamente/totalmente Bayesiano. Usualmente, ele é considerado como uma melhoria do critério de informação de desvio (DIC - *Deviance Information Criterion*) para modelos Bayesianos. Embora o DIC seja bem popular na ciência aplicada, devido a praticidade na utilização de pacotes de modelagem gráficas como BUGS, ele proporciona alguns problemas oriundos de índices que se baseiam em uma estimativa pontual (índices que não são totalmente Bayesianos). Como exemplo desses problemas podemos citar que o DIC pode produzir estimativas negativas do número efetivo de parâmetros em um modelo, além de não ser definido para modelos singulares. WAIC é totalmente bayesiano no sentido de que usa toda a distribuição a posteriori e é assintoticamente igual à validação cruzada

Bayesiana. Ao contrário do DIC, WAIC é invariante à parametrização e funciona para modelos singulares (VEHTARI; GELMAN; GABRY, 2017).

- Critério de informação e número efetivo de parâmetros.

Normalmente as medidas de precisão preditivas são referidas como critérios de informação e são normalmente definidas com base no desvio (o log da densidade preditiva dos dados, dada uma estimativa pontual do modelo ajustado multiplicado por -2, isto é, $-2\ log\ p(y \mid \hat{\theta})$).

Qualquer acurácia preditiva calculada fora da amostra normalmente é menos precisa do que a implícita predição preditiva dentro do conjunto da amostra. Em palavras mais simples, a precisão das previsões de dados futuros de um modelo ajustado geralmente será menor, na expectativa, do que a precisão das previsões do mesmo modelo para dados observados.

Lembrando que é através da precisão de predição que podemos medir o desempenho de um modelo que estamos utilizando ou simplesmente gostaríamos de comparar diferentes modelos avaliados. Não apenas para selecionar o modelo com menor erro de predição estimado ou mesmo uma média sobre os modelos candidatos como discutido por Gelman et al. (2003). Mas também para colocar modelos diferentes em uma escala comum. Mesmo considerando modelos com parametrizações completamente diferentes, podem ser usados para prever as mesmas medições.

É válido ressaltar que quando diferentes modelos têm o mesmo número de parâmetros estimados da mesma maneira, pode-se simplesmente comparar suas densidades preditivas logaritmizada do melhor ajuste, diretamente. Porém, ao comparar modelos completamente diferentes ou com número efetivo de parâmetros diferentes (por exemplo, comparando regressões logísticas ajustadas usando distribuições uniformes, *spline*, ou distribuições *a priori* de processo gaussianos), é importante fazer alguns ajustes para favorecer modelos mais parcimoniosos e evitar escolhas de modelos que naturalmente por serem mais complexos se ajustem mais facilmente aos dados, mesmo que por acaso.

- Estimando a precisão preditiva fora da amostra usando os próprios dados disponíveis.

Vários métodos foram desenvolvidos para estimar a precisão preditiva esperada usando os próprios dados disponíveis, sem precisar considerar dados futuros inacessíveis. Como não é possível estimar diretamente dados futuros até então desconhecidas $(\tilde{y}_i)$, definimos a esperança

da densidade preditiva logaritmizada fora da amostra como:

$$\begin{aligned} elpd &= \text{Esperança da Densidade Preditiva logaritmizada} \\ &= E_f(\log\ p_{post}(\tilde{y}_i)) = \int (\log\ p_{post}(\tilde{y}_i))\ f(\tilde{y}_i)\ d\tilde{y} \end{aligned}$$

em que $p_{post}(\tilde{y}_i)$ é a densidade preditiva para $\tilde{y}_i$, induzida pela distribuição a posteriori $p_{post}(\theta)$ e $f(\cdot)$ é a verdadeira distribuição do modelo que é desconhecida na prática. Por conveniência, seguimos notações matemáticas semelhantes de p_{post} introduzidas por Gelman, Hwang e Vehtari (2014) representando a distribuição a posteriori para evitar mostrar explicitamente o condicionamento de inferências sobre os dados observados y, facilitando o raciocínio para expressões mais complexas abordadas mais a frente. E vamos representar a Esperança da Densidade Preditiva Logaritmizada para uma nova observação pontual pelo seu acrônimo inglês *elpd - Expected log Predictive Density*.

Infelizmente, a expressão acima não pode ser calculada diretamente porque não conhecemos a verdadeira distribuição do modelo $f(\cdot)$. Em vez disso, se considera aproximações para contornar esse problema. Consequentemente qualquer medida de precisão preditiva baseada em métodos assintóticos será apenas uma estimativa aproximada.

Para manter a comparabilidade da aproximação com o conjunto de dados fornecido, pode-se definir uma medida de precisão preditiva para os n pontos de dados tomados um de cada vez (pontualmente):

$$\begin{aligned} elppd &= \text{Esperança da Densidade Preditiva logaritmizada Pontual} \\ &= \sum_{i=1}^{n} E_f(\log\ p_{post}(\tilde{y}_i)) \end{aligned}$$

que deve ser definido com base em alguma divisão combinada dos dados y em pontos de dados individuais y_i. A vantagem de usar uma medida pontual, em vez de trabalhar com a distribuição preditiva posterior conjunta, é que $p_{post}(\tilde{y}_i)$ está relacionado com à validação cruzada, permitindo algumas abordagens bastante gerais para a aproximação do ajuste fora do conjunto da amostra usando os dados disponíveis. Seguindo o padrão, vamos representar a Esperança da Densidade Preditiva Logaritmizada Pontualmente para um novo conjunto de dados pelo o acrônimo inglês *elppd - Expected log Pointwise Predictive Density*.

Na prática, a avaliação da precisão preditiva em um modelo ajustado, o parâmetro θ não é conhecido, e assim não podemos calcular a densidade preditiva logaritmizada $\log p(y \mid \theta)$. Por

razões óbvias normalmente em modelos Bayesianos se trabalha com a distribuição a posteriori $p_{post}(\theta) = p(\theta \mid y)$ e resumimos a precisão do modelo ajustado aos dados por:

$$\begin{aligned} lppd &= \text{Densidade Preditiva logaritmizada Pontual} \\ &= \log \prod_{i=1}^{n} p_{post}(y_i) = \sum_{i=1}^{n} \log \int p(y_i \mid \theta)\, p_{post}(\theta)\, d\theta \end{aligned} \tag{2.5}$$

novamente representamos a Densidade Preditiva Logaritmizada calculada Pontualmente pelo seu acrônimo do inglês *lppd - log pointwise predictive density*.

Utilizando métodos numéricos para calcular as posteriores do modelo, podemos atualizar a expressão anterior usando a amostragem das cadeias MCMC simuladas da $p_{post}(\theta)$, rotuladas por Gelman, Hwang e Vehtari (2014) como a Densidade Preditiva logaritmizada Pontual Computada, onde teremos o valor de θ para cada iteração θ^s, $s = 1, \ldots, S$

$$\begin{aligned} lppd \text{ computada} &= \text{Densidade Preditiva logaritmizada Pontual Computada} \\ &= \sum_{i=1}^{n} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(y_i \mid \theta^s) \right). \end{aligned} \tag{2.6}$$

Normalmente assumimos que o número de iterações S da simulação é grande o suficiente para capturar totalmente a distribuição a posteriori. Assim, sempre referimos como indistintos o valor teórico $lppd$ (equação 2.5) e o valor $lppd$ camputado (equação 2.6).

2.5.1 Critério de informação de Watanabe-Akaike (WAIC)

O Critério de Informação de Watanabe-Akaike é uma medida de precisão preditiva ajustada dentro do conjunto da amostra que fornecem estimativas aproximadamente imparciais da esperança da densidade preditiva pontual logaritmizada $elppd$ baseada na subtração do número efetivo de parâmetros (nep) e a estimativa da densidade preditiva pontual logaritmizada $lppd$.

$$\widehat{elppd} = lppd - nep.$$

Em outras palavras, o que o método de aproximação assintótico objetiva é estimar $elppd$ através de uma correção da estimativa viesada de $lppd$. Portanto, é uma abordagem Bayesiana mais completa para estimar expectativa fora da amostra (expressão 2.5), baseada na densidade preditiva a posteriori pontual logaritmizada computada corrigida para o número efetivo de parâmetros evitando o sobreajuste do modelo (*overfitting*).

Devido a importância do número efetivo de parâmetros (nep) como um fator de correção evitando o sobreajuste do modelo, é muito importante uma breve discussão sobre o assunto. Em casos padrões (onde temos um caso especial de um modelo linear normal com distribuições *a priori* não informativas como, por exemplo, uma uniforme), o nep é igual a quantidade de parâmetros a serem estimados no modelo. E assim, quanto mais parâmetros tiver o modelo (mais complexo o modelo será), maior será a precisão preditiva do mesmo. Porém, quando saímos desse universo padrão e vamos para o mundo mais prático, seguimos além dos modelos lineares com prioris uniformes. Para modelos com prioris informativas ou estrutura hierárquicas por exemplo, o número efetivo de parâmetros depende fortemente da variância dos parâmetros à nível de grupo. Normalmente, prioris informativas e estruturas hierárquicas tendem a reduzir a quantidade de sobreajuste, em comparação com o que aconteceria em mínimos quadrados simples ou estimativa de máxima verossimilhança (GELMAN; HWANG; VEHTARI, 2014).

Com relação ao critério de informação de Watanabe-Akaike, dois ajustes foram propostos na literatura. Ambos são baseados em cálculos pontuais e podem ser vistos como aproximações para validação cruzada, com base em derivações não mostradas aqui. A primeira abordagem é uma penalização p_{waic}, semelhante àquela usada para construir nep no Critério de Informação de *Deviance* ou Desvio (DIC).

$$\begin{aligned} \widehat{elppd}_{waic} &= lppd - nep \\ &= lppd - p_{waic} \end{aligned}$$

em que p_{waic} pode ser

$$p^1_{waic} = 2\sum_{i=1}^{n} \big(\log(E_{post}\, p(y_i \mid \theta)) - E_{post}(\log p(y_i \mid \theta)) \big),$$

que pode ser calculado a partir de simulações substituindo as esperanças pelas médias das S amostras desenhada pela posteriori θ^s:

$$p^1_{waic}\ \text{computado} = 2\sum_{i=1}^{n} \left(\log \left(\frac{1}{S} \sum_{s-1}^{S} p(y_i \mid \theta^s) \right) - \frac{1}{S} \sum_{s-1}^{S} \log p(y_i \mid \theta^s) \right).$$

A segunda forma, usa a variação de termos individuais da densidade preditiva logarítmica somada aos n pontos de dados. Portanto, teremos que p_{waic} da expressão 2.7 será:

$$p^2_{waic} = \sum_{i=1}^{n} Var_{post}(\log p(y_i \mid \theta)).$$

A expressão acima é mais estável do que DIC porque calcula a variância separadamente para cada ponto dos dados e então soma. A soma resulta em mais estabilidade.

Para calcular a versão através dos métodos computacionais, calcula-se a variância da densidade a posteriori preditiva logaritmizada para cada ponto dos dados y_i, ou seja, $V_{s=1}^{S} \log p(y_i \mid \theta^s)$, em que $V_{s=1}^{S}$ representa a variação da amostra $V_{s=1}^{S} a_s = \frac{1}{S-1} \sum_{s=1}^{S} (a_s - \bar{a})^2$. Somando todos os pontos de dados y_i resulta no número efetivo de parâmetros:

$$p^2_{waic} \text{ computado} = \sum_{s=1}^{S} V_{s=1}^{S} \Big(\log p(y_i \mid \theta^s) \Big).$$

Portanto podemos utilizar qualquer versão de p^1_{waic} ou p^2_{waic} como uma correção de viés. Gelman, Hwang e Vehtari (2014) compararam ambos e recomendam, em estudos práticos, que utilize p^2_{waic} porque sua expansão em série tem uma semelhança mais próxima com a expansão em série para validação cruzada deixando um de fora (LOO-CV) e na prática, os resultados de p^2_{waic} e LOO-CV parecem estarem mais próximos um do outro, aparentemente.

Tal como acontece com AIC e DIC, definimos WAIC multiplicando por -2, para deixar na escala do desvio (*deviance*) e facilitar a comparação:

$$\widehat{elppd}_{waic} = -2\, lppd + 2\, p_{waic}.$$

Na definição original de Watanabe, WAIC é o negativo da densidade preditiva pontual logarítmica média (assumindo a previsão de um único ponto para um novo dado) e, portanto, é dividido por n e não tem o fator 2. Seguimos a expressão sugerida por Gelman, Hwang e Vehtari (2014) para que os AIC, DIC e outras medidas de desvio sejam comparáveis entre elas. Para mais detalhes sobre a comparação entre p^1_{waic} e p^2_{waic}, bem como informações das propriedades desejáveis entre AIC, DIC e WAIC, sugiro a leitura de Watanabe e Opper (2010), Gelman, Hwang e Vehtari (2014) e Vehtari, Gelman e Gabry (2017).

2.5.2 Método de Validação Cruzada deixando um de fora (LOO-CV)

Na validação cruzada Bayesiana, os dados são repetidamente particionados em um conjunto de treinamento y_{treino} e um conjunto de validação y_{valid}. O modelo é ajustado para y_{treino} (produzindo assim uma distribuição a posteriori $p_{treino}(\theta) = p(\theta \mid y_{treino})$), o qual é avaliado usando uma estimativa da densidade preditiva logarítmica com os dados de validação, $\log p_{treino}(y_{valid}) = \log \int p_{pred}(y_{valid} \mid \theta)\ p_{treino}(\theta)\ d\theta$. Assumindo que a distribuição a posteriori $p(\theta \mid y_{treino})$ é resumida por S simulações da amostragem θ^s, calculamos a densidade preditiva logarítmica como $\log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_{valid} \mid \theta^s)\right)$.

Para simplificar, restringiremos nossa atenção aqui à validação cruzada deixando um de fora (LOO-CV do inglês *leave-one-out cross-validation*), sendo um caso especial n partições em que cada conjunto de validação representa um único ponto dos dados. Realizar a análise para cada um dos n pontos dos dados, resultará em n inferências diferentes $p_{post(-i)}$, cada uma resumida por S simulações posteriores, θ^{is}.

A estimativa Bayesiana LOO-CV de ajuste preditivo fora da amostra será:

$$lppd_{loo-cv} = \sum_{i=1}^{n} \log p_{post(-i)}(y_i), \text{ calculado como } \sum_{i=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i \mid \theta^{is})\right).$$

Cada previsão é condicionada a $m-1$ pontos nos dados, o que causa subestimação do ajuste preditivo. Para n grande, a diferença é desprezível, porém para n pequeno (ou usando validação cruzada *k-fold*), podemos usar uma correção de viés de primeira ordem b, estimando quantas previsões melhores seriam obtidas se o condicionamento fosse feito em n pontos nos dados (BURMAN, 1989):

$$b = lppd - \overline{lppd}_{-i},$$

em que,

$$\overline{lppd}_{-i} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \log p_{post(-i)}(y_j), \text{ calculado como } \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_j \mid \theta^{is})\right).$$

O LOO-CV Bayesiano corrigido pelo viés (denominado de *cloo-cv*) é então:

$$lppd_{cloo-cv} = lppd_{loo-cv} + b.$$

A correção de polarização b raramente é usada, pois geralmente é pequena, mas a incluímos para ser completa.

A validação cruzada é semelhante com WAIC no sentido que os dados sejam divididos em partes separadas, de preferência condicionalmente independentes. Isso representa uma limitação da abordagem quando aplicada a modelos estruturados. Além disso, a validação cruzada pode ser computacionalmente cara, exceto em configurações onde os atalhos estão disponíveis para aproximar as distribuições $p_{post(-i)}$ sem ter que reajustar o modelo a cada vez.

Sob algumas condições, diferentes critérios de informação mostraram ser assintoticamente iguais à validação cruzada *leave-one-out* (com $n \to \infty$, a correção de viés pode ser ignorada). A validação cruzada Bayesiana também funciona com modelos singulares[2], e o LOO-CV Bayesiano provou ser assintoticamente igual a WAIC (WATANABE; OPPER, 2010). Para n finito, há diferença, pois LOO-CV condiciona as densidades preditivas a posteriori em $n-1$ pontos dos dados. Essas diferenças podem ser aparentes para pequenos n ou em modelos hierárquicos.

[2] Se a matriz de informação de Fisher for positiva definida para todo o vetor de parâmetro θ, então o modelo estatístico correspondente é dito regular; caso contrário, o modelo estatístico é considerado singular por definição.

3 CONSIDERAÇÔES FINAIS

A aquicultura é uma ciência multidisciplinar complexa, de grande relevância socioeconômica para o Brasil e fascinante como área investigativa científica, cujos problemas ou desafios estão longe de serem totalmente esgotados ou exauridos em uma simples Tese. Quatro anos de investigações e aprendizado, são relativamente pouco tempo para erigir uma Tese (levantar uma hipótese), desenvolver uma teoria com base na hipótese, criar uma tecnologia, validá-la e expectar, no caso das ciências aplicadas, que toda informação gerada cause um impacto positivo e alcance a sociedade. Porém nesse período, foi possível investigar algo de grande relevância para o setor aquícola, em particular para a administração das fazendas de camarão, que é o crescimento do *Litopenaeus vannamei* ao longo de uma produção real. Com base em dados da produção de uma fazenda de camarão cinza no Nordeste Brasileiro foi possível modelar o crescimento dos animais estocados em viveiros comerciais com baixas densidades de estocagem. No entanto, antes de definir e validar um modelo promissor que representasse o desenvolvimento dos crustáceos ao longo do tempo, foi detectado pela primeira vez na literatura a possível subestimativa na estimação dos parâmetros em modelos não lineares sigmoides utilizados usualmente na aquicultura.

Portanto, antes de modelar o crescimento foi necessário identificar os motivos das subestimações dos parâmetros nos modelos e propor um método de correções para esse viés. Dessa forma, baseado nas curvas sigmoides não lineares do primeiro capítulo, concluímos que os dados classificados como dados incompletos ou limitados, são restritos a observações abaixo do ponto de inflexão da curva. Ou seja, os pesos dos animais (g) ao longo do tempo observado nas fazendas de produções são restritos em até 12g ou 18g. Isso ocorre porque os maiores tamanhos de animais amostrados (tamanho final de abate) são limitados ao tamanho exigido pelo mercado-alvo. Portanto, mostramos que as inferências de parâmetros para modelos não lineares usando as abordagens frequentistas a partir de dados em fazendas comerciais de camarão ou mesmo em condições de laboratório podem estar viesadas.

Os modelos hierárquicos bayesianos foram sugeridos como um método para resolver este problema, bem como uma poderosa ferramenta de gerenciamento dentro de uma fazenda de camarão, melhorando a eficiência da produção e para pesquisas científicas, uma vez que a nova abordagem fornece índices comparativos mais sensíveis. A principal vantagem técnica

deste método é que as estimativas dos parâmetros são abordadas como efeitos aleatórios, apresentando assim densidades de probabilidades, diferente da estatística clássica. Os parâmetros da curva não linear têm um significado biológico interessante para a aquicultura. Nesta abordagem, sua estimativa é mais precisa e permite o uso de testes de hipóteses comparativos. Além disso, é possível realizar análises em qualquer nível da hierarquia (nível do tanque ou nível do ciclo de produção) capturando as características gerais de todos os tanques ou viveiros da fazenda, e diferenças individuais em cada ciclo de produção, sendo muito interessante para a gestão de uma fazenda.

Baseado no método proposto, foi possível fazer a modelagem do crescimento do camarão cinza ao longo do tempo. Entre vários modelos não lineares ajustados (Morgan-Mercer-Flodin, Michaelis-Menten, Weibull, von Bertalanffy, Gompertz e a equação de crescimento Logístico) a equação de crescimento de Weibull se destacou como a que melhor descreve o fenômeno, considerando a estrutura Hierárquica Bayesiana para modelar o crescimento do camarão branco do Pacífico (*Litopenaeus vannamei*). Embora o diagnóstico de ajuste do modelo indicou baixa especificação ao nível da população, ao nível do grupo viveiro e ao nível do ciclo produtivo obtiveram excelentes resultados. Concluindo que o modelo foi considerado bem ajustado para tais níveis hierárquicos. Assim, o instrumento é profícuo já que os níveis hierárquicos desses grupos são interessantes na análise do manejo dentro das fazendas de camarão.

Os resultados também mostraram maior sensibilidade da nova ferramenta na detecção de diferenças entre possíveis tratamentos comparativos. Portanto, este instrumento pode ser utilizado na melhoria de processos e produtos dentro de um ambiente comercial voltado para a moderna Indústria 4.0 na aquicultura.

REFERÊNCIAS

AKAIKE, H. Information theory and an extension of the maximum likelihood principle,[w:] proceedings of the 2nd international symposium on information, bn petrow, f. **Czaki, Akademiai Kiado, Budapest**, 1973.

BARBOSA, A. B. R. **História e evolução da carcinicultura no Rio Grande do Norte**. Dissertação (B.S. thesis) — Universidade Federal do Rio Grande do Norte, 2022.

BENZIE, J. H. Population genetic structure in penaeid prawns. **Aquaculture Research**, Wiley Online Library, v. 31, n. 1, p. 95–119, 2000.

BERNARDO, J.; SMITH, A. **Bayesian theory**. [S.l.]: Wiley, New York, 1994.

BETANCOURT, M. Identifying the optimal integration time in hamiltonian monte carlo. **arXiv preprint arXiv:1601.00225**, 2016.

BETANCOURT, M. A conceptual introduction to hamiltonian monte carlo." arxiv preprint. **URL https://arxiv. org/pdf/1701.02434. pdf**, 2017.

BETANCOURT, M.; BYRNE, S.; GIROLAMI, M. Optimizing the integrator step size for hamiltonian monte carlo. **arXiv preprint arXiv:1411.6669**, 2014.

BISHOP, C. M. **Pattern recognition and machine learning**. [S.l.]: springer, 2006.

BOX, G. E.; TIAO, G. C. **Bayesian inference in statistical analysis**. [S.l.]: John Wiley & Sons, 1992. v. 40.

BOYD, C. E.; THUNJAI, T. Concentrations of major ions in waters of inland shrimp farms in china, ecuador, thailand, and the united states. **Journal of the World Aquaculture Society**, Wiley Online Library, v. 34, n. 4, p. 524–532, 2003.

BURMAN, P. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. **Biometrika**, Oxford University Press, v. 76, n. 3, p. 503–514, 1989.

CASTRO, A.; CAVALCANTI-MATA, M.; DUARTE, M. Avaliação do sabor de filés de camarão (litopenaeus vannamei) submetidos a diferentes condições de congelamento e armazenamento. In: **congresso brasileiro de ciências e tecnologia de alimentos**. [S.l.: s.n.], 2004. v. 21.

CHALONER, K. Elicitation of prior distributions. **Bayesian biostatistics**, Marcel Dekker New York, p. 141–156, 1996.

CHONG-ROBLES, J. et al. Osmoregulation pattern and salinity tolerance of the white shrimp litopenaeus vannamei (boone, 1931) during post-embryonic development. **Aquaculture**, Elsevier, v. 422, p. 261–267, 2014.

DAVIS, D. A.; SAMOCHA, T. M.; BOYD, C. **Acclimating Pacific white shrimp, Litopenaeus vannamei, to inland, low-salinity waters**. [S.l.]: Southern regional aquaculture center Stoneville, Mississippi, 2004.

DRAPER, D. Model uncertainty yes, discrete model averaging maybe. **Statistical Science**, v. 14, p. 405–409, 1999.

DUANE, S. et al. Hybrid monte carlo. **Physics letters B**, Elsevier, v. 195, n. 2, p. 216–222, 1987.

EPIFANI, I. et al. Case-deletion importance sampling estimators: Central limit theorems and related results. **Electronic Journal of Statistics**, The Institute of Mathematical Statistics and the Bernoulli Society, v. 2, p. 774–806, 2008.

FAO. **The State of World Fisheries and Aquaculture 2020: Sustainability in action**. Rome: Food and Agriculture Organization of the United Nations, 2020.

FRENCH, S. Group consensus probability distributions: A critical survey in bayesian statistics. **Bayesian statistics**, v. 2, 1985.

GARTHWAITE, P. H.; KADANE, J. B.; O'HAGAN, A. Statistical methods for eliciting probability distributions. **Journal of the American Statistical Association**, Taylor & Francis, v. 100, n. 470, p. 680–701, 2005.

GELFAND, A. E. Model determination using sampling-based methods. **Markov chain Monte Carlo in practice**, London, p. 145–161, 1996.

GELFAND, A. E.; DEY, D. K.; CHANG, H. **Model determination using predictive distributions with implementation via sampling-based methods**. [S.l.], 1992.

GELFAND, A. E.; MALLICK, B. K.; DEY, D. K. Modeling expert opinion arising as a partial probabilistic specification. **Journal of the American Statistical Association**, Taylor & Francis, v. 90, n. 430, p. 598–604, 1995.

GELFAND, A. E.; SMITH, A. F. Sampling-based approaches to calculating marginal densities. **Journal of the American statistical association**, Taylor & Francis Group, v. 85, n. 410, p. 398–409, 1990.

GELMAN, A. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). **Bayesian analysis**, International Society for Bayesian Analysis, v. 1, n. 3, p. 515–534, 2006.

GELMAN, A. et al. Hierarchical models. **Bayesian data analysis**, Chapman and Hall/CRC New York, p. 120–160, 2003.

GELMAN, A. et al. **Bayesian data analysis**. [S.l.]: Chapman and Hall/CRC, 2013.

GELMAN, A.; HWANG, J.; VEHTARI, A. Understanding predictive information criteria for bayesian models. **Statistics and computing**, Springer, v. 24, n. 6, p. 997–1016, 2014.

GELMAN, A.; MENG, X.-L.; STERN, H. Posterior predictive assessment of model fitness via realized discrepancies. **Statistica sinica**, JSTOR, p. 733–760, 1996.

GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, n. 6, p. 721–741, 1984.

GENEST, C.; SCHERVISH, M. J. Modeling expert judgments for bayesian updating. **The Annals of Statistics**, JSTOR, p. 1198–1212, 1985.

HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. Oxford University Press, 1970.

HOETING, J. A. et al. Bayesian model averaging: a tutorial. **Statistical science**, JSTOR, p. 382–401, 1999.

HOFFMAN, M. D.; GELMAN, A. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. **Journal of Machine Learning Research**, v. 15, n. 1, p. 1593–1623, 2014.

IBGE. **Instituto Brasileiro de geografia e Estatística**. [S.l.], 2020. v. 48. Accessed: 2022-04-12. Disponível em: <https://sidra.ibge.gov.br/pesquisa/ppm/quadros/brasil/2020>.

JAYNES, E. T. **Probability theory: the logic of science**. [S.l.]: Washington University St. Louis, MO, 1996.

JOHNSON, S. R. et al. Methods to elicit beliefs for bayesian priors: a systematic review. **Journal of clinical epidemiology**, Elsevier, v. 63, n. 4, p. 355–369, 2010.

KADANE, J.; WOLFSON, L. J. Experiences in elicitation: [read before the royal statistical society at a meeting on'elicitation 'on wednesday, april 16th, 1997, the president, professor afm smith in the chair]. **Journal of the Royal Statistical Society: Series D (The Statistician)**, Wiley Online Library, v. 47, n. 1, p. 3–19, 1998.

LARSON, H. **Introduction to Probability Theory and Statistical Inference**. [S.l.]: John Wiley and Sons, 1982.

LEIMKUHLER, B.; REICH, S. **Simulating hamiltonian dynamics**. [S.l.]: Cambridge university press, 2004. v. 14.

LINDLEY, D. Reconciliation of probability distributions. **Operations Research**, INFORMS, v. 31, n. 5, p. 866–880, 1983.

LINDLEY, D. V. Reconciliation of discrete probability distributions. **Bayesian statistics**, North Holland Amsterdam, v. 2, n. 375-390, 1985.

MACKAY, D. J.; KAY, D. J. M. **Information theory, inference and learning algorithms**. [S.l.]: Cambridge university press, 2003.

METROPOLIS, N. et al. Equation of state calculations by fast computing machines. **The journal of chemical physics**, AIP, v. 21, n. 6, p. 1087–1092, 1953.

NASH, C. **The history of aquaculture**. [S.l.]: John Wiley & Sons, 2010.

NAYLOR, J. C.; SMITH, A. F. Applications of a method for the efficient computation of posterior distributions. **Journal of the Royal Statistical Society: Series C (Applied Statistics)**, Wiley Online Library, v. 31, n. 3, p. 214–225, 1982.

NEAL, R. Bayesian learning for neural networks [phd thesis]. **Toronto, Ontario, Canada: Department of Computer Science, University of Toronto**, 1995.

NEAL, R. M. et al. Mcmc using hamiltonian dynamics. **Handbook of markov chain monte carlo**, v. 2, n. 11, p. 2, 2011.

NUNES, A. O cultivo do camarão litopenaeus vannamei em águas oligohalinas. **Panorama da Aquicultura**, v. 11, n. 66, p. 26–35, 2001.

NUNES, H. R. Acompanhamento da implantação de um laboratório de produção de pós-larvas de camarão marinho. Florianópolis, 2003.

OAKLEY, J. E.; O'HAGAN, A. Uncertainty in prior elicitations: a nonparametric approach. **Biometrika**, Oxford University Press, v. 94, n. 2, p. 427–441, 2007.

O'HAGAN, A. et al. **Uncertain judgements: eliciting experts' probabilities**. [S.l.]: John Wiley & Sons, 2006.

PAULINO, C. D. M.; TURKMAN, M. A. A.; MURTEIRA, B. **Estatística bayesiana**. [S.l.: s.n.], 2003.

PERUGGIA, M. On the variability of case-deletion importance sampling weights in the bayesian linear model. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 92, n. 437, p. 199–207, 1997.

PHILLIPS, L.; WISBEY, S. The elicitation of judgmental probability distributions from groups of experts: a description of the methodology and records of seven formal elicitation sessions held in 1991 and 1992. **Nirex UK, Didcot, UK: Report nss/r282**, 1993.

ROBERT, C.; CASELLA, G. **Monte Carlo statistical methods**. [S.l.]: Springer, 2004.

ROBERT, C.; CASELLA, G. **Monte Carlo statistical methods**. [S.l.]: Springer Science & Business Media, 2013.

ROY, L. A.; DAVIS, D. A. Requirements for the culture of the pacific white shrimp, litopenaeus vannamei, reared in low salinity waters: water modification and nutritional strategies for improving production. **Avances en Nutrición Acuicola**, 2010.

RUBIN, D. B. Bayesianly justifiable and relevant frequency calculations for the applies statistician. **The Annals of Statistics**, JSTOR, p. 1151–1172, 1984.

SAMOCHA, T. M. et al. Design and operation of super intensive, biofloc-dominated systems for indoor production of the pacific white shrimp, litopenaeus vannamei–the texas a&m agrilife research experience. **Louisiana: The World Aquaculture Society. 368p**, 2017.

SCHOBER, J. Pesquisa impulsiona produção de camarões em viveiros e mercado de trabalho regional. **Ciência e Cultura**, Sociedade Brasileira para o Progresso da Ciência, v. 54, n. 1, p. 10–11, 2002.

STAN, D. Stan: A c++ library for probability and sampling, version 2.14.0. **Online: http://mc-stan. org**, 2017.

STONE, M. An asymptotic equivalence of choice of model by cross-validation and akaike's criterion. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 39, n. 1, p. 44–47, 1977.

VEHTARI, A.; GELMAN, A.; GABRY, J. Practical bayesian model evaluation using leave-one-out cross-validation and waic. **Statistics and computing**, Springer, v. 27, n. 5, p. 1413–1432, 2017.

WATANABE, S.; OPPER, M. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. **Journal of machine learning research**, v. 11, n. 12, 2010.

WEST, M.; CROSSE, J. Modelling probabilistic agent opinion. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 54, n. 1, p. 285–299, 1992.

SEGUNDA PARTE

ARTIGO 1 - ***Evidence of parameters underestimation from nonlinear growth models for data classified as limited***

Redigido conforme as normas da revista *Computers and Electronics in Agriculture* (versão final já publicada).

Computers and Electronics in Agriculture 200 (2022) 107196

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Evidence of parameters underestimation from nonlinear growth models for data classified as limited

Carlos Antônio Zarzar [a,b,*], Edilson Marcelino Silva [b], Tales Jesus Fernandes [b], Izabela Regina Cardoso De Oliveira [b]

[a] Universidade Federal do Oeste do Pará (UFOPA), Av. Maj. Francisco Mariano CEP: 68220-000, Cidade Alta, Monte Alegre, Pará, Brazil
[b] Universidade Federal de Lavras (UFLA), 3037 CEP: 37200-900, Lavras, Minas Gerais, Brazil

ARTICLE INFO

Keywords:
Sigmoid models
Shrimp farm
Penaeus vannamei
Growth modeling
Bayesian hierarchical

ABSTRACT

Some growth data in aquaculture have peculiar characteristics that generate consequences in the analysis and modeling. They are usually incomplete or limited, as classified in this article. This means data are restricted to a few observations and often are limited to observations below the curve's inflection point due to economic interests in farm settings, or due to limitation of physical space in controlled research laboratories, for example. This possibly causes under and/or overestimation in the inference of nonlinear models. Through shrimp growth simulations from the Michaelis–Menten curve, the limited data were synthesized with threshold observation up to the first 7, 13, 18, 36, and 82 weeks. Seven sigmoid growth functions (Logistic, Gompertz, von Bertalanffy, Richard, Weibull, Morgan–Mercer–Flodin, and the own Michaelis–Menten growth) were fitted to respective limited data, in order to assess the research hypothesis. Taking the scenarios with incompleteness in the first 7, 13 and 18 weeks, the parameters of all growth curves modeled under a frequentist approach were underestimated. Thus, we propose a correction for this possible problem through a hierarchical Bayesian approach. Real data from shrimp farming in northeastern Brazil were used to compare it with the traditional frequentist approach employed. The sensitivity in detecting outstanding treatment (pond or batch level hierarchy) can make the new method a powerful management tool in animal production, and also in trials designed for scientific research.

1. Introduction

Shrimp farming has been eminent worldwide for its growth and economic performance of USD 69.3 billion (9.4 million tonnes) in 2018 (Fao, 2020). In Brazil, its production was 63.2 thousand tons in 2020, generating economics for the country with a Gross Production Value of USD 282.14 million in 2020 (IBGE, 2020), and estimated substantial growth of 23.8% in 2021. Thus, it is natural the scientific growth and efforts in several areas of interest to meet the growing demand for this food around the world.

Some aquaculture researchers have recently focused on commercial rather than experimental settings (Yu et al., 2006; Yu and Leung, 2010; Ruiz-Velazco et al., 2010; Estrada-Pérez et al., 2016). This approach presumably brings better inferential results for predicting growth purposes, because the samples are taken directly from the investigated universe (e.g fish/shrimp farms), instead of being generalized from a controlled experiment. Such experiments are featured by a high control degree and uniformity of dependent variables, which is unlike meeting in a commercial shrimp farm for example. Therefore, an empirical shrimp growth curve from a commercial farm may diverge from an experimentally derived shrimp growth curve under a laboratory setting (Yu et al., 2006).

A problem that arises while using company data in predicting growth curves is that they are usually incomplete. That means the largest animal sizes sampled (final size) on a farm are restricted to the size required by the target market. Thus, the information for modeling a whole growth curve is incomplete and is usually limited to observations below the curve's inflection point due to economic interests.

Modeling a whole growth curve is important because it enables us to infer about parameters that have a biological meaning (e.g. maximum relative growth rate, first maturation size, maximum asymptotic size) which are interesting for farming (Fernandes et al., 2015). However, incomplete data can lead to wrong estimates for some of these parameters. Thus, under or overestimated parameters pose difficulty in the growth curves comparison among pond, seasons, broodstock, larvae

* Corresponding author at: Universidade Federal do Oeste do Pará (UFOPA), Av. Maj. Francisco Mariano CEP: 68220-000, Cidade Alta, Monte Alegre, Pará, Brazil.
E-mail address: carlos.zarzar@ufopa.edu.br (C.A. Zarzar).

https://doi.org/10.1016/j.compag.2022.107196
Received 15 February 2022; Received in revised form 29 June 2022; Accepted 1 July 2022
Available online 14 July 2022

from the different breeding stock origin, cultivation technology, or any comparison of interest in aquaculture.

The Bayesian statistical approach has shown promise for solving incomplete data problems in recent years (Zhou et al., 2020; Luo and Kareem, 2020; Shi and Tong, 2020). This is due to the fact that the Bayesian theory considers the Prior information in addition to the data information which can enables better inferences, estimates, and precision. Furthermore from the Posterior distribution, it is possible to update the belief about the Prior Information, and based on the new evidence, the Posterior distribution can be updated bringing the perspective of continuous learning.

In general, exact posterior inference in Bayesian requires an intractable analytic calculus. However, Bayesian statistics has been advanced in approximate statistical inference methods that resort to computational numerical integration. In this paper, we use the No-U-Turn Sampler (NUTS) algorithm (Hoffman and Gelman, 2014), derived from the Hamiltonian Monte Carlo (HMC) (Neal, 2012) that belongs to the Markov chain Monte Carlo (MCMC) methods family. Besides being a very fast algorithm, NUTS is an adaptive, that is, some necessary hyperparameters for efficient sampling are automatically set. It cunningly does this, during the burn-up period, taking advantage of discarded samples to set them.

1.1. About MCMC sampling algorithms and its parameters

Markov Chain Monte Carlo (MCMC) methods comprise a family of algorithms (e.g. Metropolis–Hastings, Gibbs Sampling, Hamiltonian Monte Carlo, Slice sampling, No-U-Turn) for random sampling from a posterior probability distribution (Metropolis and Ulam, 1949; Neal, 2012; Hoffman and Gelman, 2014; Betancourt, 2017). It is a numerical integration method widely used in Bayesian inference, when the integration is very complex or when there is no analytical solution. In a simplified way, we can see the MCMC sampler as a temporal process composed of iterations that samples the unknown parametric space at random (a process known as a random walk), constructing a Markov chain. Each chain, in turn, must be homogeneous, irreducible, and ergodic with a stationary distribution, as an inherent characteristic. To obtain desired approximate numerical integration, a relatively large number of iterations is necessary, and more than one chain is recommended. As the chain starts from a random starting point and takes time to reach the convergence, we call this period the burn-in (or warm-up) period. We discard them so that it does not cause bias in the Bayesian inference.

A important observation in the MCMC method is that normally is introduced an independency between interactions so that the chain reaches convergence faster and consequently finds the target posterior distribution faster too. Although this dependence on MCMC samples does not interfere with the validity of the Bayesian inference, it does affect the efficiency of the sampler. Correlated MCMC samples require more iterations to produce the same level of Monte Carlo error for an estimate. Therefore, after sampling, it is necessary to remove this autocorrelation from the interactions. The thinning removes that dependency. Hence, we can declare these important MCMC parameters in the model tuning process as hyperparameters that always need to be set on a case-by-case basis (Metropolis and Ulam, 1949; Neal, 1993, 2012; Hoffman and Gelman, 2014; VanDerwerken and Schmidler, 2017; Betancourt, 2017).

1.2. BayesIan hierarchical models

Bayesian hierarchical models have been one of the main tools in statistical decision theory (Berger, 1985) and this form of modeling is increasingly common in operations research, management science (Mauritzen, 2020), and agricultural research (Li et al., 2022). Given that aquaculture data have a natural hierarchical structure (cultivation cycles, pond and farm levels), the manager can benefit by capturing the overall characteristics of all tanks on the farm and also allowing individual differences of each cycle productions (Gelman et al., 2013; Murphy, 2012). The generalized perspective is interesting for the manager in making more general decisions for the company, but it is also interesting to understand the peculiarities of each crop to seek improvements for the business. Thereby, the hierarchical structure implemented in this Bayesian nonlinear proposal is suggested in order to correct the parameter estimation bias caused by limited data.

1.3. Research goals and contributions

The goals of this research are: (1) to evaluate the problem of parameter under and/or overestimation in the growth curve modeling under a frequentist approach and (2) to propose a method based on hierarchical Bayesian modeling to be applied to incomplete data from commercial settings or from other limiting conditions as ones in small research laboratories.

2. Materials and methods

This paper is motivated by a real problem that is common in shrimp farms: incomplete data may cause problems in parameter estimation while modeling nonlinear growth curves under a frequentist approach. Such a problem is investigated in a simulation study, which is described in Section 2.1. We considered different sigmoid curves for quantified the sensitivity of incomplete data on the bias of the parameter estimation. As a second phase of this research, we propose a method based on the Bayesian hierarchical model approach using the Hamiltonian Monte Carlo method (HMC) from Markov Chain Monte Carlo (MCMC) to solve the problem.

2.1. The simulation study

In order to investigate underestimation from shrimp farming data, incomplete data from the generalized Michaelis–Menten equation was simulated by restricting the animal size from the first weekly biometrics observations. The generalized Michaelis–Menten equation was chosen since it provides a flexible model for different animal species which is able of describing a sigmoidal growth behavior with its variable inflection point (Lopez et al., 2000; Tagliafico et al., 2018). Its asymmetric characteristic provides a high growth rate at the beginning of cultivation and a gradual decrease after the inflection point in accordance with the shrimp growth in farming. The generalized Michaelis–Menten equation is [see Lopez et al. (2000) for more details]:

$$w(t) = \frac{w_0\beta^\kappa + \alpha t^\kappa}{\beta^\kappa + t^\kappa} + \varepsilon$$

where $w(t)$ is the weight as a function of time, the β parameter can be interpreted as the time (week unit) that the animal will reach the arithmetic mean of the weight at the beginning (the weight intercept when $t = 0$) and at the end of the animal's life (the weight when $t = \infty$). We can notice when $t = \beta$, we will have $w(t = \beta) = \frac{w_0\beta^\kappa + \alpha\beta^\kappa}{\beta^\kappa + \beta^\kappa} = \frac{w_0+\alpha}{2}$. If $w_0 = 0$ the Michaelis–Menten function is in the enzyme kinetics equation form (Michaelis and Menten, 1913) with the time replacing substrate concentration. In this condition, it can be assert that the β parameter is the time when half-maximal weight is achieved like $w_0 = 0 \Rightarrow w(t = \beta) = \frac{\alpha}{2}$. The κ parameter is strongly associated with the inflection point coordinates in a growth model. The $\beta > 0$ and $\kappa > 0$ must be positive since the growth is strictly increasing and to ensure that the rate of variation weight (derived from the function) has a maximum point, κ must be greater than one ($\kappa > 1$) (Lopez et al., 2000).

The w_0 and α parameters are respectively the animal weight at the time equal to zero and time tending to infinity (asymptotic theoretical weight), that is, the weight at the end of the animal's life (w_∞). t denotes animal age in weeks ($t > 0$) and ε is the random error with

Table 1
Nonlinear equations (sigmoidal growth curves) fitted to simulated incomplete data for shrimp farming (*Litopenaeus vannamei*) in order to investigate parameters underestimation.

Name	Equation
Logistic growth	$w(t) = \frac{\alpha}{1 + e^{\kappa(\beta - t)}}$
Gompertz	$w(t) = \alpha \cdot e^{-e^{\kappa(t-\beta)}}$
von Bertalanffy	$w(t) = \alpha \left[1 - e^{-\kappa(t-\beta)}\right]^3$
Richard	$w(t) = \alpha \left[1 + (\gamma - 1)e^{-\kappa(t-\beta)}\right]^{\frac{1}{1-\gamma}}$
Weibull	$w(t) = \alpha \left(1 - e^{-\beta t^{\gamma}}\right)$
Morgan–Mercer–Flodin	$w(t) = \alpha - \frac{\alpha - \beta}{1 + (\kappa\ t)^{\gamma}}$
Michaelis–Menten growth	$w(t) = \frac{w_0 \beta^{\kappa} + \alpha t^{\kappa}}{\beta^{\kappa} + t^{\kappa}}$

normal distribution $N(\mu, \sigma)$ with zero mean ($\mu = 0$) and the standard deviation sigma [$\sigma = \tau \cdot w(t)$] as a function of average weight weighted by (τ) parameter in order to simulate the heterogeneity of variance. For data simulation, the parameters of the Michaelis–Menten growth equation assumed the following values: $w_0 = 0.2g$, $\alpha = 90.58g$, $\beta = 47.7$, $\kappa = 1.2$, $\varepsilon \sim N(\mu = 0, \sigma = \tau \cdot w(t))$ with $\tau = 0.08$, defined based on a prior information described in Section 2.2.1.

Limited/incomplete data were simulated with thresholds of 7, 13, 18, 36, and 82 weekly biometric observations. Such values were based on the average real production cycle of the shrimp (*Litopenaeus vannamei*) in ponds (usually 7, 13, and 18 weeks in northeastern Brazil) and the time reached for the expected asymptotic weight (around 36 and 82 weeks, see Section 2.2.1). Seven nonlinear growth curves were fitted in this first phase of research, through the frequentist statistical approach (nonlinear least squares - Gauss–Newton algorithm). Math expressions are presented in Table 1.

2.2. The proposed method to deal with the underestimation problem

The solution proposed in this research is the Bayesian hierarchical model estimated using the No-U-Turn (NUTS) algorithm (Hoffman and Gelman, 2014) derived from the Hamiltonian Monte Carlo method (HMC) (Betancourt, 2017), which belongs to Markov Chain Monte Carlo (MCMC) family (Neal, 2012).

2.2.1. Elicitation of prior distributions

The methods for the elicitation of prior information in an appropriate manner are very important for the application of Bayesian inferences (Moala and Penha, 2016). In this paper, the prior density distributions were obtained by maximum entropy method, a meta-analysis of existing studies given in Table 2, and also information from the expert production manager.

- Maximum entropy method

In some cases, we have no prior knowledge about the parameter information. So a frequent solution in statistical studies is to use the principle of indifference. The principle of maximum entropy tells us how to extend the principle of indifference to such cases. One crucial point for understanding the entropy information theory is the Shannon entropy ($H(x) = -\sum_{k=1}^{k} p_k\ log(p_k)$ in discrete case and $\int_{-\infty}^{\infty} f_x(x)\ log[f_x(x)]dx$ for continuous cases) (Shannon, 1948). According to the theory, to maximize entropy means to obtain the maximum uncertainties about the studied phenomenon (Jaynes, 2003; Park and Bera, 2009). In most cases to optimize the restricted entropy function, the Lagrange multiplier method is used. Therefore, even without knowing the prior distribution $P(x)$, it is possible to find it making the fewest assumptions (Singh et al., 1986) since known some statistical moment about $P(x)$, such as the mean or variance, is known.

According to Singh et al. (1986), we can generate some prior distributions based on the principle of maximum entropy. The knowledge about the first and second moments around the mean of the variable can generate *Normal* (β parameter) or *Lognormal* distribution (α parameter) depending on whether the variable undergoes any logarithmic transformation and if it assumes asymmetry in the distribution, respectively. If all known information is a range of possible values of the parametric space domain, the method leads to a *Uniform* (κ parameter) distribution prior. Finally, if the knowledge about the parameter is the first moment and the asymmetry is verified, the prior distribution is derived from a *Gamma* (τ parameter) distribution.

- Meta-analysis and expert information

The meta-analysis was carried out from several papers with the species *L. vannamei* wild (fishing). The largest total length of animals sampled or the maximum total length estimated in the surveys was organized in Table 2 as well as the maximum weight when this information was available. When it was not available, it was estimated using the weight-to-length relationship given by $W(L) = 0.00000283L^{3.22}$ according to Ramos-Cruz (2011) (W is the weight in grams to be estimated and L is the total length of white shrimp).

The parameter α on the proposed model (Section 2.2.2) can be interpreted as the maximum asymptotic weight that the animal can reach at the end of its life. Its prior distribution was elucidated as a *lognormal* with parameter $\mu_{\alpha} = 4.3729$ and $\sigma_{\alpha} = 0.3137$ for $E[X] = 83.27$ and $SD[X] = 2{\times}13.39 = 26.78$ obtained directly from previous scientific studies or indirectly through the relation weight and length estimated in some specific searches (Table 2). It was chosen to use twice the standard deviation found in the meta-analysis in order not to be too restrictive since it is an estimate and therefore there are uncertainties for the estimated value. We assume a positive skewness in the *lognormal* distribution, not discarding aquaculture information and weighing its estimates at lower values than in fishing for this parameter.

The parameter κ is associated with the time (t) in which the weight gain is maximum, that is, it is strongly associated with the weight changes rate at the inflection point (curve slope) in a growth model. Therefore, no estimates of this Michaelis–Menten curve parameter for shrimp growth have been reported in the literature. But since the inflection point can be interpreted biologically as the maturation age, it was possible to extract some information from the farm manager, from the maturity researches of *L. vannamei* shrimp and from the simulation previously carried out. Therefore, owing to the little information associated with this parameter, we were cautious with this *Prior* and opted for a uniform distribution within a extreme range of $0.1 \leq \kappa \leq 5.5$ [$Uniform(\delta_{\kappa} = 0.1, \lambda_{\kappa} = 5.5)$].

The w_0 parameter was fixed to 0.2 grams due to information from the shrimp larviculture company considering the hatching weight of the animals and low variation of this value. It represents the weight at the beginning of the animal's life.

The β parameter can be associate to the time when half-maximal weight is achieved in the growth model. Considering the meta-analysis performed for the maximum theoretical asymptotic weight, its half will be equal to $41.635g$. Based on the expert production manager the animal can reach this weight from one year (48 weeks) to one year and four months (64 weeks) on average at 58 weeks. Thereby, it was defined as *Normal* as the *Prior* distribution with $\mu_{\beta} = 58.0$ and $\sigma_{\beta} = 10.0$ parameters.

Finally, for the τ parameter associated to the variability of the shrimp average weight (σ), a parametrization in function of average weight was used ($\sigma = \tau \cdot \mu$) with the *Gamma* probability distribution for τ *Priori* information ($\gamma = 1; \theta = 1$). The choice was based on the maximum entropy, the simulation study and the restriction of these parameters to be strictly positive. The prior parameter values considered in this paper are summarized in Table 3. Also, they were also used to specify initial values for all model parameters in order to obtain faster convergence of the chains and lower computational costs.

Table 2
The meta-analysis of previous studies for the *Prior* information about the α parameter.

Weight asymptotic (g)	Length asymptotic (mm)	Sample size	Region	Country	Note[a]	References
87.90	213.00	5,104	Gulf of Tehuantepec	Mexico	Observed information	Ramos-Cruz (2011)
78.64[d]	205.00	3,955	Gulf of Tehuantepec	Mexico	Estimated information	Cervantes-Hernández et al. (2017)
113.91[d]	230.00	NA	Eastern Pacific	Mexico and Peru	FAO catalog	Holthuis (1990)
72.63[d]	200.00	NA	Sinaloa and Sonora	Mexico	Observed information	Lluch (1974)
80.00	NA	NA	NA	NA	Personal communication	Tian et al. (1993)
77.20[d]	210.00	30 adult females	Sinaloa	Mexico	Observed information	Hernández-Covarrubias et al. (2012)
72.63[d]	200.00	NA	Sinaloa	Mexico	NA	Chávez (1973)
83.27/13.39	209.14/9.60	MEAN[b] /SD[c]				

[a]Observation indicating whether the total length was from a larger observed sample or if the maximum total length was estimated by the article.
[b]$E[X]$ - The average value asymptotic in the meta-analysis.
[c]$SD[X]$ - The standard deviation of asymptotic value in the meta-analysis.
[d]Weight values estimated through the relationship $W(L) = 0.00000283L^{3.22}$ according to Ramos-Cruz (2011).
NA - Not Available.

Table 3
Prior information for the parameters of the Bayesian hierarchical model.

Bayesian hierarchical parameters	The prior probability distribution	Values of parameters
w	$Normal(\mu, \sigma)$	μ, $\sigma = \tau \cdot \mu$ to be estimated
τ	$Gamma(\gamma, \theta)$	$\gamma = 1$ and $\theta = 1$
α	$Lognormal(\mu_\alpha, \sigma_\alpha)$	$\mu_\alpha = 4.3729$ and $\sigma_\alpha = 0.3137$
κ	$Uniform(\delta_\kappa, \lambda_\kappa)$	$\delta_\kappa = 0.1$ and $\lambda_\kappa = 5.5$
β	$Normal(\mu_\beta, \sigma_\beta)$	$\mu_\beta = 58.0$ and $\sigma_\beta = 10.0$
w_0	Fixed parameter	$w_0 = 0.2$

2.2.2. The model

Many regression models have been studied as appropriate functional forms for a shrimp growth curve by Tian and Dong (2006), Tian et al. (1993) as the linear models, polynomial, log reciprocal, von Bertalanffy, Gompertz, Logistic, Exponential, including machine learning models as Artificial Neural Network (ANN) (Yu et al., 2006). Nonetheless, in this research, we adopted the Michaelis–Menten growth equation due to the properties reported in Section 2.1.

Thus, the Bayesian hierarchical nonlinear model proposed can be expressed as follows:

$$w_{ij} \sim Normal(\mu_{ij}, \sigma_{ij})$$

$$\mu_{ij} = \frac{w_0 \beta_{ij}^{\kappa_{ij}} + \alpha_{ij}\, t^{\kappa_{ij}}}{\beta_{ij}^{\kappa_{ij}} + t^{\kappa_{ij}}} \quad ; \quad \sigma_i j = \tau_{ij} \cdot \mu_{ij}$$

$$\tau_{ij} \sim Gamma(\gamma, \theta)$$

$$\alpha_{ij} \sim Lognormal(\mu_\alpha, \sigma_\alpha)$$

$$\kappa_{ij} \sim Uniform(\delta_\kappa, \lambda_\kappa)$$

$$\beta_{ij} \sim Normal(\mu_\beta, \sigma_\beta)$$

where w_{ij} is the shrimp average weight in each biometrics over time, i is the indexer for each pond (tank) with $i = 1, \ldots, I$; j if the indexer for each cycle production with $j = 1, \ldots, J$, μ_{ij} is the expected value of average weight over the time fitted through the proposed model, σ_{ij} is the standard deviation of the average weight (μ_{ij}) along the time, τ_{ij} is a perturbation coefficient over the weight mean, γ; θ are *Gamma* distribution parameters, as $\mu_\alpha, \sigma_\alpha$ are *Lognormal* distribution parameters, μ_β; σ_β are *Normal* distributions parameters; $\delta_\kappa, \lambda_\kappa$ are *Uniform* distribution parameters. Finally α_{ij}, κ_{ij}, w_0 and β_{ij} are the Michaelis–Menten model parameters.

In the proposed method we consider heterogeneity of variance, therefore it takes into account the weight variability over time, through the expression $w_{ij} \sim Normal(\mu_{ij}, \sigma_{ij} = \tau_{ij} \cdot \mu_{ij})$. This means that the standard deviation is a function of the expected weight, thus as the animal weight increases over time the variance increases proportionally. In future research, we intend to improve the model by considering other methods and expressions for modeling heteroscedasticity, but to simplify this proposal and following the scope of the research, a parsimonious model was defined.

2.2.3. Proposed method assumptions

Regardless of the model fitted through the proposed method, there are assumptions to be made. (i) regarding the parameters, we usually assume that the parameter α is the maximum weight that the individual can achieve in life, which means that it is a characteristic inherent to the specie that can vary from one individual to another, therefore, it is a theoretical asymptotic weight (w_∞) to be estimated. As it is a parameter that reflects the genetic potential of the animal, in favorable cultivation conditions, we assume that all individuals can express it, and naturally, they will be able to reach this theoretical asymptotic weight. (ii) If for any reason a stress agent starts during farming (it is understood as a stressor agent: lack of food, high concentrations of nitrogen compounds, temperature, scarce oxygen, etc.), obviously, it will influence the relative growth rate. However, if it is not long-lasting and if it is removed as soon as possible from cultivation, the animal may return to its potential growth previously depict (according to the theory of compensatory growth) (Gallardo-Collí et al., 2020; Mohanty, 2015; Wasielesky Jr. et al., 2013; Ali et al., 2003). (iii) However, if these stress agents in culture persist, then the animal will not be able to print all its genetic potential (either to some nutritional deficiency, biological dysfunction, physical conditions of space, deterioration of water quality, etc.), and then this can lead to lower estimates for α parameter or even the animal's death.

2.2.4. Computational numerical method

The model was fitted using the Stan probabilistic programming language (Stan Development Team, 2020) with R interface (R Core Team, 2020), which implements HMC with No-U-Turn (NUTS) algorithm. The sample was run with 6,000 iterations in each of 4 chains. The 3,000 (50%) first samples of each chain of the simulation were discarded as a burn-in period. Since convergence through HMC is faster than other methods in the MCMC family, this burn-in amount was enough as shown through diagnostic analyses in the appendix. In order to avoid autocorrelation trends during the simulated samples, it was defined as the thinning values 4th every sample, and this configuration was enough as shown through graphical analysis in the appendix. Beware, the thinning removes that dependency, nevertheless, while it reduces the final autocorrelation of the samples, and the total number of samples saved.

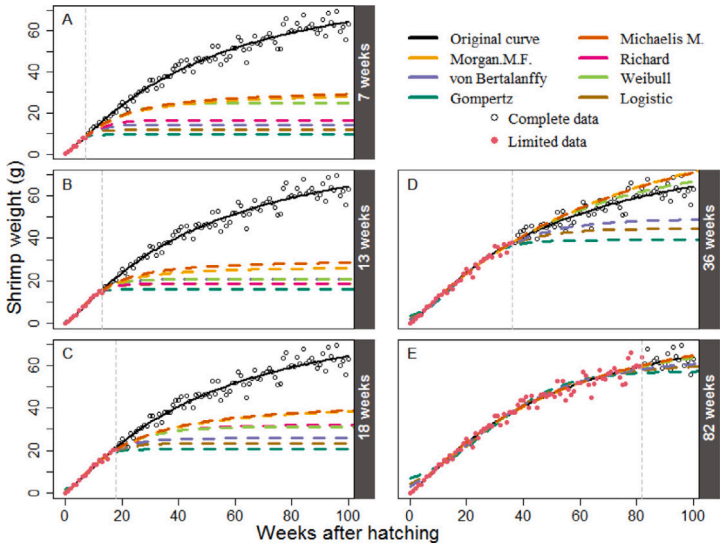


Fig. 1. Different nonlinear models fitted to simulated incomplete data of shrimp (*Litopenaeus vannamei*) weight under a frequentist approach and considering five scenarios: incompleteness at (A) 7, (B) 13, (C) 18, (D) 36, and (E) 82 weeks of biometrics measurement.

2.3. Real data and cultivation characteristic

The data was provided by a commercial shrimp farming entity culturing white shrimp (*L. vannamei*) in Northeast of Brazil, which operated 55 ponds between 1.5 and 19.3 hectares throughout the 2017 and 2018 year. For this research, weekly sampling data were gathered only from 40 grow-out ponds at this farm, covering 205 production cycles during that range time. The production cycles lasted between 5 weeks (38 days) and 15 weeks (102 days). Post-larvae (PL20) juvenile shrimps at 0.2–0.5 g were stocked (density between 2 to 30 ind. m^{-2}) and then harvested at an average weight and standard deviation (sd) of 8.43 ± 0.72 g. The dataset consists of 1820 observations.

3. Results

3.1. Simulation result

Fig. 1 shows simulated data for shrimp (*L. vannamei*) weight (grams) over 100 weeks from the generalized Michaelis–Menten equation. Taking the first 7, 13, and 18 weeks, the parameters of all growth curves fitted under a frequentist approach were underestimated (Fig. 1 A, B, and C). For incomplete data at 36 and 82 weeks, convergence was not achieved for the Richard equation (Fig. 1 D, and E). Although the Michaelis–Menten, Morgan–Mercer–Flodin, and Weibull models are close to the original curve for 36 weeks threshold of incomplete data (Fig. 1 D), there is a very high degree of uncertainty about the α parameter which is noticed through the estimate of the standard error, $\alpha_{36} = 118.32 \pm 54.67$ g (mean $\pm$ standard error for Michaelis–Menten); $\alpha_{36} = 123.26 \pm 77.74$ g (Morgan–Mercer–Flodin) and $\alpha_{36} = 76.38 \pm 30.58$ g (Weibull growth). For the Logistic, Gompertz, and von Bertalanffy models, the α_{36} parameters were below 50 grams (minimum 39.15 and maximum 49.03 grams).

Obviously, in general, the best results were for samples with data from 82 weeks threshold (Fig. 1 E). In general, the larger the sample, the closer it gets to the population, allowing reliable inferences. The alpha parameter from the Logistic model ($\alpha_{82} = 57.33 \pm 1.04$ g), von Bertalanffy ($\alpha_{82} = 62.82 \pm 1.53$ g), and Gompertz ($\alpha_{82} = 60.61 \pm 1.31$ g) were underestimated in relation to the true value $\alpha = 90.58$ g. Furthermore, such models showed poor fits for the extreme values (the Initial weight at the limit of the week tending to zero and the theoretical asymptotic weight at the limit of the week tending to infinity) (Fig. 1E). An intermediate but still underestimated result for 82 weeks threshold was found for the Weibull model ($\alpha_{82} = 70.57 \pm 4.70$ g). Finally, the closest estimates to the true value ($\alpha = 90.58$ g) were from the Morgan–Mercer–Flodin ($\alpha_{82} = 91.01 \pm 11.04$ g) and the Michaelis–Menten ($\alpha_{82} = 91.08 \pm 8.73$ g) curves (Fig. 1 E).

Usually, the alpha parameter (α) is a common parameter among most nonlinear growth models. Comparing the other parameters between the different models is more difficult because they have distinct values and biological meanings for each function. However, we can compare the estimated parameters of the Michaelis–Menten function with the true values that generated the simulated data. The beta parameter estimates were (mean$\pm$standard error) $\beta_7 = 14.8350 \pm 27.1809$, $\beta_{13} = 12.153 \pm 4.550$, $\beta_{18} = 19.5653 \pm 7.9090$, $\beta_{36} = 70.678 \pm 47.286$, and $\beta_{82} = 48.10696 \pm 8.19881$ for the limits of 7, 13, 18, 36, and 82 weeks respectively. It always remained below the true value $\beta = 47.7$ for the 18-week threshold or less. At week 36 the estimated range includes the true value, although the low precision level. The best result was for the 86 weeks with estimates close to the true value and a better level of accuracy.

The kappa parameter for all thresholds studied were $\kappa_7 = 1.3441 \pm 0.4337$, $\kappa_{13} = 1.537 \pm 0.235$, $\kappa_{18} = 1.3410 \pm 0.1698$, $\kappa_{36} = 1.115 \pm 0.128$, and $\kappa_{82} = 1.19703 \pm 0.09208$. The range estimates contained the true value of $\kappa = 1.2$. However, obviously, the accuracy improves as the number of observations increases.

3.2. Chains convergence diagnostic

Before any inference about the posterior distribution of the model, it is very important to check chain convergences. In summary, the convergence diagnostic result (Appendix) indicated that there was no evidence of non-convergence in the chains estimated by the MCMC method. Evidence of non-autocorrelation was observed. There was no divergence in the sampling after the burn-in.

3.3. BayesIan inference on the estimated parameters

Once the quality of convergence is guaranteed regardless of the sampling method used in numerical integration, inference becomes more reliable. This means that the parametric space of the unknown posterior distribution has been well explored and numerically quantified by the defined algorithm. Therefore, considering the fitted nonlinear Bayesian

Table 4
Descriptive statistics for estimates of α, β, κ, and τ parameters, obtained through the proposed and usual methods. 205 cycles production and for 40 ponds from a shrimp farm.

		Parameters	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
Bayesian method	Cycle	$Alpha\,(\alpha)$	55.15	94.25	103.29	101.77	111.53	136.95
		$Beta\,(\beta)$	34.32	48.84	52.02	51.93	54.63	63.47
		$Kappa\,(\kappa)$	0.99	1.28	1.35	1.37	1.46	1.89
		$Tau\,(\tau)$	0.09	0.28	0.39	0.43	0.50	1.32
	Tank	$Alpha\,(\alpha)$	59.75	84.19	94.13	95.40	106.46	129.99
		$Beta\,(\beta)$	41.41	50.06	53.69	52.70	55.58	60.82
		$Kappa\,(\kappa)$	1.16	1.26	1.32	1.33	1.39	1.56
		$Tau\,(\tau)$	0.29	0.43	0.53	0.56	0.66	1.97
Usual method	Cycle	$Alpha\,(\alpha)$	8.24	10.85	14.28	51.86	22.37	3460.20
		$Beta\,(\beta)$	3.63	5.10	6.74	14.43	11.95	321.24
		$Kappa\,(\kappa)$	1.16	1.71	2.04	2.29	2.73	5.05
	Tank	$Alpha\,(\alpha)$	6.02	8.30	10.10	10.66	11.58	20.86
		$Beta\,(\beta)$	3.20	4.44	5.38	5.74	6.42	11.17
		$Kappa\,(\kappa)$	1.65	2.05	2.29	2.57	2.77	5.85

hierarchical model, for each 205 shrimp production cycles we will have a sigmoidal curve obtained through the parameter estimates. If we consider the pond as the hierarchical level of interest, we will have 40 growth curves, one for each tank on the farm. In summary, the descriptive statistics of estimates for the α, β, κ, and τ parameters are shown in Table 4 for all production cycles and also for the pond hierarchical level. In addition, the same table shows estimates for the same parameters, estimated using the traditional frequentist method. Nonetheless, for this last approach, 58 production cycles found no solutions or had serious problems in the fit of the model or did not converge, and therefore, were not estimated.

Based on the Michaelis–Menten function of growth fitted to the limited or incomplete data by the Bayesian proposed method, the lowest estimated value for the mean of the α parameter was for the production cycle number 141 at tank 6 $\alpha_{6,141} = 55.15g$, against $\alpha_{6,141} = 24.93g$ by usual frequentist method for the same pond and cycle production. The maximum value estimated for alpha average was $\alpha_{13,24} = 136.95g$ by the proposed method for the production cycle number 24 at tank 13, against $\alpha_{13,24} = 61.02g$ by frequentist method. For the β parameter, the minimum value estimated for the mean by the proposed method was $\beta_{22,194} = 34.32$ weeks corresponding to the production cycle number 194 at pond 22, and the maximum value was $\beta_{6,141} = 63.47$ weeks corresponding to the 141 cycle number, against $\beta_{22,194} = 42.85$ and $\beta_{6,141} = 31.89$ by traditional method for the respective pond and cycle. Regarding the κ parameter estimated by the proposed method, the lowest value found for the mean was $\kappa_{2,128} = 0.99$, and the highest was $\kappa_{22,194} = 1.89$, against $\kappa_{2,128} = 1.55$ and $\kappa_{22,194} = 1.95$ by traditional method, for the respective ponds and production cycles.

Table 5 shows statistics of the posterior distribution of parameters for three different trials (cycle productions: 31, 120, and 176). The results consist of mean, standard deviation, credible interval, the Effective sample size (n_{eff}), and the potential scale reduction statistic ($\hat{R}$).

Finally, Fig. 2 shows the parameter estimates for each pond shrimp farming with 80% and 95% credible intervals in order to compare the pond quality production.

Table 5
The posterior distribution estimated by the Bayesian hierarchical model for three cycle productions: 31 (pond 1), 120 (pond 4) and 176 (pond 28).

Parameters	Mean	SD[a]	Cred. Interval		n_{eff}[b]	$\hat{R}$[c]
[cycle number]			2.5%	97.5%		
Alpha[31] (α_{31})	88.02	22.98	51.63	140.69	2735	1.00
Alpha[120] (α_{120})	104.48	25.39	63.33	161.35	3067	1.00
Alpha[176] (α_{176})	85.56	20.59	52.42	131.23	2839	1.00
Kappa[31] (κ_{31})	1.30	0.15	1.02	1.61	2910	1.00
Kappa[120] (κ_{120})	1.47	0.11	1.25	1.68	2974	1.00
Kappa[176] (κ_{176})	1.33	0.12	1.11	1.56	2862	1.00
Beta[31] (β_{31})	55.69	9.49	37.33	75.08	2522	1.00
Beta[120] (β_{120})	51.45	9.06	34.63	70.39	3106	1.00
Beta[176] (β_{176})	56.44	9.00	39.37	74.65	2665	1.00
Tau[31] (τ_{31})	0.64	0.18	0.39	1.07	2820	1.00
Tau[120] (τ_{120})	0.25	0.09	0.15	0.48	2770	1.00
Tau[176] (τ_{176})	0.32	0.11	0.18	0.59	2795	1.00

[a]Standard deviation of estimated parameter.
[b]Effective sample size.
[c]Potential scale reduction statistic.

4. Discussion

4.1. Main research hypothesis

An important aspect of modeling the growth of living organisms is that the estimated parameters cannot be biased, otherwise incorrect conclusions will be drawn. The problem while using the company data or experimental design, in predicting growth curves is that they are usually incomplete/limited. Thus not all points over the curve are sampled, and observations are restricted to below the curve's inflection point (smaller animals). The hypothesis of this paper is that this restriction usually leads to under or overestimated the parameters of nonlinear growth models. The Bayesian approach applied to nonlinear models with informative Priors probabilities functions provides better parameter estimates being a powerful method to solve this kind of problem (Zhou et al., 2020; Luo and Kareem, 2020; Shi and Tong, 2020; Agyeman et al., 2022).

4.2. Results that support the research hypothesis

Our simulation results support strongly this hypothesis. Furthermore, it indicated a sensitivity in the underestimation of parameters up to the 36th week of the trial (Fig. 1 D). Studies with other animals, from a few observations, also indicated the same behavior, collaborating with the hypothesis of this research. Salles et al. (2020) compared Bayesian and frequentist approaches for growth curves in Santa Inês sheep (the Gompertz growth model comprising the period of up to 210 days). They pointed out that when using smaller samples with limited data to 120 days, the α parameter estimated by the frequentist approach was 18% smaller than the one obtained through the Bayesian approach, underestimating the asymptotic weight. Although the authors did not point out the real reason for this underestimation in detail, this research sheds light on the problem by specifying the possible reason and quantifying the sensitivity in the shrimp farming application. Also, we proposed a method that takes into account the hierarchical structure of the model using the HMC-NUTS algorithm, more efficient and more reliable in terms of the Markov chains diagnostic.

The results with real data from shrimp farming (see Table 4) also indicated bias in the parameter estimates, contributing to the research hypothesis.

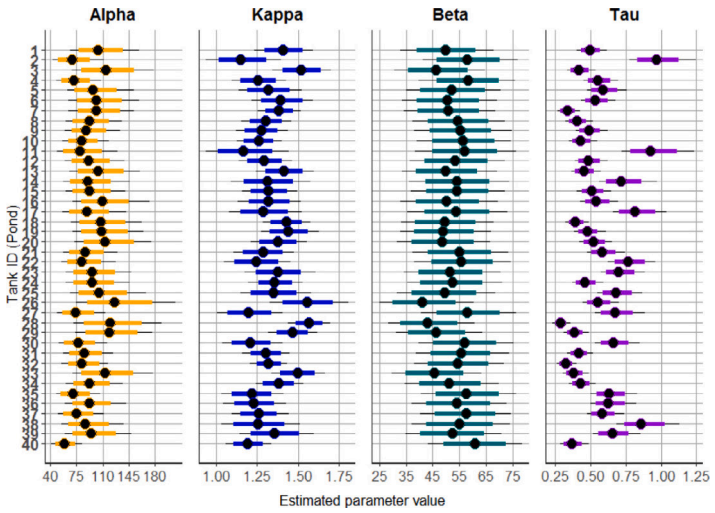


Fig. 2. Estimated parameters of each tank with 80% and 95% credible interval for 40 pond of shrimp (*Litopenaeus vannamei*) farming modeled with the Bayesian hierarchical approach proposed with generalized Michaelis Menten as the nonlinear statistic model.

4.3. Comparison with the literature results

Many journal publications in the aquaculture field have presented estimated values for the parameters from nonlinear growth models quite controversial. For the theoretical asymptotic weight (α parameter) presented by Hernandez-Llamas and Ratkowsky (2004), for example, they diverge completely from the values estimated for fisheries data for the same species (Table 2), regardless of the nonlinear sigmoidal model assumed by the research. Tian et al. (1993) estimated the α parameter for von Bertalanffy equation and for Gompertz equation around $\alpha = 29.86$ g and $\alpha = 29.78$ g respectively. Yu and Leung (2010) estimated from a Logistic equation over frequentist inference and Bayesian hierarchical approach $\alpha = 43.60 \pm 0.61$ g and $\alpha = 32.10 \pm 2.09$ g respectively. More recently, Aragón-Noriega et al. (2017) estimated $\alpha = 39.55$ (38.38 – 40.76 confidence interval) for a Gompertz equations, claiming to be the best model fitted to the data. Yu et al. (2006) estimated for Logistic and von Bertalanffy equations values around $\alpha = 42.27$ g and $\alpha = 31.93$ g respectively (weight values estimated through the relationship $W(L) = 0.00000283L^{3.22}$ according to Ramos-Cruz (2011)). Similar results were found in our simulation when data were restricted to 7 up to 36 weeks. In contrast, the wild shrimp was estimated asymptotic weight average $\alpha = 83 \pm 13.39$ g (Table 2).

The aforementioned authors disregard the effect of limited data on the underestimation. They argue that estimates for the α parameter are correlated to other cultivation variables such as temperature, and water quality parameters, mainly the stocking density. That is environmental and management factors under which the shrimp is submitted. It is believed that they may provide lower alpha values since the animal grows in an artificial environment, unlike wild environments.

They are completely different environments and difficult to compare, wild animals do not have the "comfort" and conditions that aquaculture provides although. The benefits of control over water quality variables, feed availability, and balanced nutritional composition for the species of different ages and different production methods (extensive, intensive, etc.) allow the animal to print all its genetic potential and consequently reach theoretical asymptotic weight as high as wild shrimp. Moreover, In the wild environment, the animals have concerns about predators, inconsistent feeding, and a heavy dependence on environmental conditions. All these facts support our hypothesis.

The hypothesis of the article shed light on a problem that has not been pointed out. The findings can lead to major consequences, not only in shrimp farming but in any agricultural production that shares the characteristic of limited or incomplete data. We have not ruled out the hypothesis defended by the authors until then, but we have raised a new hypothesis and we suggest an alternative methodology. After all, our proposed method does not ignore the possibility of finding values estimated much lower than expected by the species caused by poor environmental and/or poor management conditions as shown in the results. Likewise, it corrects the underestimation through the prior information given by the Bayesian theory.

4.4. Sensitivity of the proposed method

The great advantage of the nonlinear function used to model growth is the biological interpretation of the parameters. The parameter estimates can be used as a management tool in a shrimp farm as well as a comparative index of each productive cycle and also between the ponds. Probably such parameters as comparative indexes estimated through the proposed method are more sensitive in identifying better performances and efficient trials than the traditional methods used, both in aquaculture farms management and scientific research.

Results that collaborate with this argument were the production cycles 31, 120, and 176 (Table 5), which presented very similar growth results in the dataset (final weight of 8.5 g in 10 weeks of cultivation). Although they were produced under different conditions, as they started at different times (March, June, and July), with different initial densities (5, 6.5, and 5.7 ind. m^{-2}), different pond sizes (12, 7.7, e 3.5 hectares), a different Feed Conversion Factor (FCA) (0.54, 0.64 and 0.79), survival of 68.0%, 39.0% and 42.0% respectively. Traditional analyses are unlikely to capture differences between these treatments (as well as an experimental design with few or no repetitions). However, taking the proposed approach advantages and using parameter estimates as an analysis tool, one can distinguish and identify the superiority between those treatments. Even perform a hypothesis test for each parameter, since according to the Bayesian principle, they are random parameters and follow a distribution of probability.

Table 5, batch 120 has the desired characteristics for farm management considering the Michaelis–Menten equation. A higher estimate of α, a lower value for the parameter β, a higher value for the parameter κ, and a low value for τ. According to the method, we clearly perceive that cultivation 120 distinguishes from cultivations 31 and 176.

5. Conclusion

The problem of parameter underestimation in nonlinear growth models was identified and emphasized. The data classified as incomplete or limited are restricted to observations below the curve's inflection point. This is because the largest animal sizes sampled (final size) on a shrimp farm are restricted to the size required by the target market. Therefore, we show that the parameter inferences for nonlinear models using the frequentist approaches from such data may be biased.

The Bayesian hierarchical models were suggested as a method to address this problem. Furthermore, we concluded that it can be a powerful investigative tool for shrimp farms and also for scientific research, since the new approach provides a more sensitive comparative index.

Possibly such a method can be applied to any research field of interest (or any animal) whose growth data may be subject to incompleteness and potentially underestimation of the model parameters. So the suggestion for future research is to carry out modeling with different growth curves based on the proposed method, such as Logistic growth function, Gompertz, von Bertalanffy, Richard, Weibull growth function, Morgan-Mercer-Flodin. In addition, it is recommended to test the method for other animals as fish, sheep, chickens, pigs, goats, cows, and other animals with production interest. Finally, other more complex proposals to take into account the variance heterogeneity are recommended, so that the model is improved and accurately reflects the phenomenon under study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgments

The authors thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We are grateful for the contribution of Mr. Eduardo de Freitas D. Antona and Sylvio Araujo de Mattos, director and president of SECOM Aquicultura Industria e Comércio, Mr. Marcelo Soares and Gustavo Novaes, productive technicians at the shrimp farm, Dr. Antonio Calixto Silva da Rocha, attorney. In addition, Mr. José Washington Bezerra businessman, took pictures of the largest specimens on his farm, supporting the hypothesis presented in this paper. C. A. Zarzar thanks everyone who are part of the Universidade Federal do Oeste do Pará (UFOPA), especially the Monte Alegre campus (Aquaculture Engineering Course) and the graduate program on Estatística e Experimentação Agropecuária at the Federal University de Lavras (UFLA) for full support during this research.

Appendix A. Diagnostic analysis method

A.1. Algorithm analysis, HMC-NUTS hyperparameters, and chains diagnosis

- The trace plot and Potential scale reduction factor ($\hat{R}$)

The algorithm's efficiency in exploring the unknown parametric space and the MCMC chains' health after reaching convergence, was evaluated both by appropriate index and through graphical analyses. The time series plot of the Markov chains known as *traceplot* is a useful diagnostic plot. It is expected that after the convergence of the chains, the graph of each MCMC chain for each parameter estimated in the model, will be stationary representing a time series of white noise (see appendix). Another tool used in this research was the potential scale reduction factor ($\hat{R}$) presented by Gelman and Rubin (1992), Gelman et al. (2013). It is a diagnostic index that attempts to flag situations where the MCMC algorithm has failed to converge. As a rule, it is expected that $\hat{R}$ values for all parameters are less than 1.1. It measures the ratio of the average variance of draws within each chain to the variance of the pooled draws across chains; if all chains are at equilibrium, these will be the same, and $\hat{R}$ will be close to 1. If the chains have not converged to a common distribution, the $\hat{R}$ statistic will be greater than 1 (see Gelman et al. (2013) for more details about this index).

- Autocorrelation plot and The effective sample size (ESS)

After the burn-in period, evidence of non-autocorrelation of chains was evaluated by graphical analysis and some specific indexes as N_{eff}/N. For serial correlation analysis of the MCMC chains, graphical analysis is usually used. The effective sample size (N_{eff}) measures the amount by which autocorrelation in samples increases uncertainty (standard errors) relative to an independent sample. If there is autocorrelation, the effective sample size (N_{eff}) will be smaller than the total number of iterations (N). It is useful to calculate the N_{eff}/N ratio. The larger the ratio of N_eff to N, it is the better (see Gelman et al. (2013) for more details). One should be concerned about a ratio less than 0.1.

- Specific diagnostics of the HMC sampler and the NUTS algorithm

The great differential of the HMC method with the NUTS algorithm is the gain of numerous additional tools for diagnostics that indicate that the sampler is breaking and, thus, not sampling from the posterior distribution. The three main HMC-NUTS specific diagnoses are listed: divergent transitions, maximum tree-depth, and Bayesian fraction of missing information (see Hoffman and Gelman (2014), Betancourt (2016b,a, 2017) for more details).

The technical details of the HMC-NUST algorithm may not be appropriate to cover here (for this purpose Betancourt and Stein (2011), Neal (2012), Betancourt and Girolami (2013), Hoffman and Gelman (2014), Betancourt (2017) are recommended), but a general understanding is necessary to interpret the diagnostic tools correctly. In a very simplified way, computational statistics inspired by mechanical physics developed the HMC algorithm with the proposal to approximate the sampler as a particle moving without friction on the posterior distribution surface. For this purpose, it was necessary to add an auxiliary momentum variable that simulates the kinetic and potential energy of the particle on the surface, and a discrete approximation of a continuous function when integrating (*leapfrog* integration), simulating the elapsed time of the particle at each iteration. If the step sizes of *leapfrog* integration are too large, the discrete approximation does not work and the approximation will be poor between the true posterior distribution and the sampled distribution, indicating divergent transitions. If there are too many divergent transitions, then the sampler is not drawing samples from the entire posterior, and inferences will be biased. Therefore it is important to have no or few divergent transitions detected by the Stan software. The target average proposal acceptance probability (*adapt_delta*) used to determine the step size hyperparameter during warm-up automatically was set to $adapt\_delta = 0.99$.

One of the hyperparameters used by NUTS that are automatically set by Stan is the tree-depth during the warm-up iterations period. For the NUTS algorithm, the tree-depth plot is an important diagnostic. If the sampler is often hitting the maximum number of steps, it means that

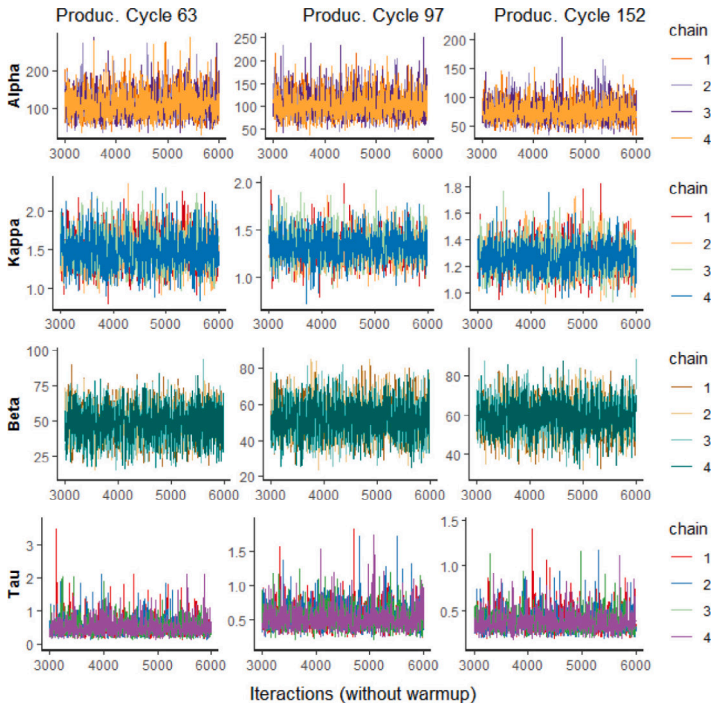


Fig. 3. Traceplot over the interactions without warm-up period for the four Markov chain for Alpha (α), Kappa (κ), Beta (β) and Tau (τ) parameter estimates for the production cycle number 63 (pond 27), 97 (pond 34) and 152 (pond 28).

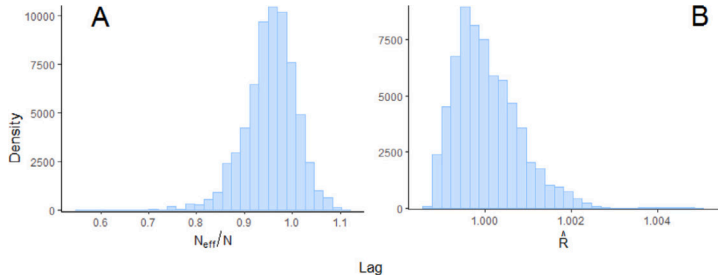


Fig. 4. (A) The histogram of the N_{eff}/N ratio draws within a Markov Chain. The effective sample size (N_{eff}) by the total sample size (N). (B) The potential scale reduction statistic ($\hat{R}$) index.

the optimal number of steps to take in each iteration is higher than the maximum. Taking many steps may be a sign of poor adaptation, perhaps due to targeting a very high acceptance rate, or may simply indicate a difficult posterior from which to sample. HMC will select a better parameter value when the parameter space bends back on itself or when the number of steps specified by max tree-depth is reached. By default, the max tree-depth is 10. A too-small maximum tree-depth only affects algorithm efficiency, exploration is slower and the autocorrelation is higher (effective sample size lower) than if the maximum tree-depth were set higher. Usually, a graphical analysis of the tree-depth hyperparameter set throughout the war-up period is sufficient as a diagnosis.

Regarding the Bayesian fraction of missing information, as mentioned, the HMC sampler augments approximation to the target posterior by adding fictitious momentum or, equivalently, the choice of a kinetic energy function. The Hamiltonian function that describes the total energy is based on mechanical statistics, which will not be detailed here (see Betancourt (2017) for more details). But in practice, the Bayesian fraction of missing information is a criterion that readily estimates using the history of energies in the Hamiltonian Markov chain. The most relevant to emphasize for this research is the energy diagnostic quantifies mainly the heaviness of the posterior distribution tails, and can be analyzed through the energy diagnostic plot (appendix) that shows overlaid histograms of the energy transition density and marginal energy distribution.

Appendix B. Results for diagnostic analysis

Chains convergence diagnostic

- The trace plot diagnostic

The time series plot of the Markov chains known as the *traceplot* is a useful diagnostic plot. It shows the evolution of the parameter vector over the iterations of the 4 Markov Chains for monitoring convergence. In Fig. 3 are the healthy traceplot of the estimated parameters Alpha (α), Kappa (κ), Beta (β), and Tau (τ) for the Cycle Productions numbers 63, 97 and 152, respectively, presented as an example.

- The Potential scale reduction factor ($\hat{R}$)

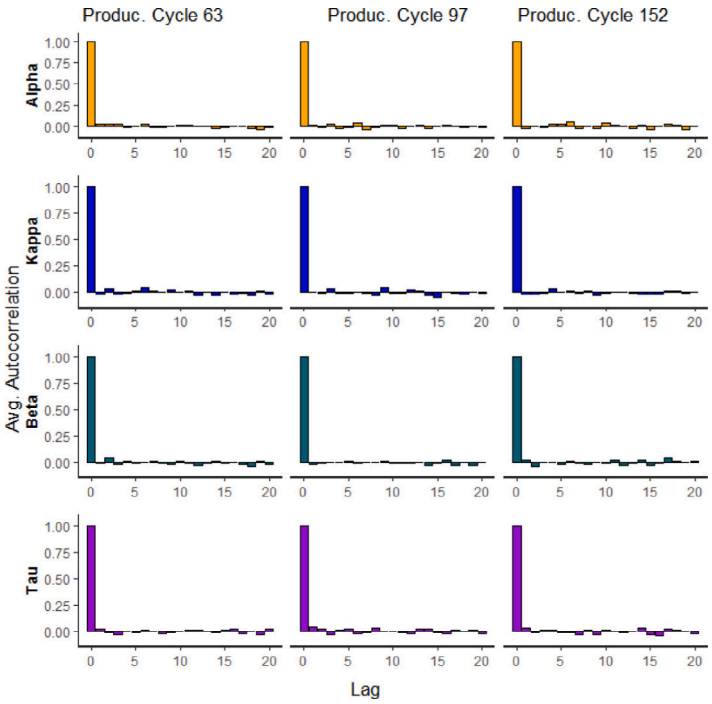


Fig. 5. Autocorrelation plot drawn within the four Markov chain for Alpha (α), Kappa (κ), Beta (β) and Tau (τ) parameter estimates for the production cycle number 63 (pond 27), 97 (pond 34) and 152 (pond 28).

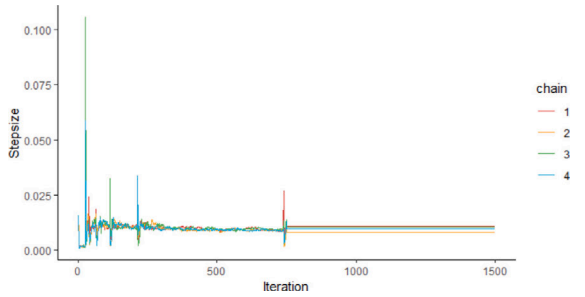


Fig. 6. Hyperparameter Stepsize converging to the optimal during the iteration warm-up period.

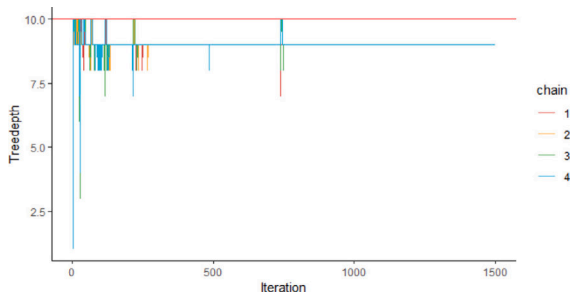


Fig. 7. Hyperparameter Treedepth below the maximum value in the iteration warm-up period converging to the optimal set value.

A power tool for diagnostic Bayesian models is the potential scale reduction statistic ($\hat{R}$). In our results we obtained an $\hat{R}$ average of 1.000052 ± 0.000749 (Fig. 4 B).

- N_{eff}/N ratio and Autocorrelation plot (correlogram)

The average N_{eff}/N (effective sample size/total number of iterations) ratio in the fitted model was 0.9534 ± 0.0543 (Fig. 4 A), which means that both the model (including the parametrization) and the particular MCMC algorithm used in this study are appropriate. It can also be observed that each parameter estimation from the posterior distribution drawn within the Markov chain was independent, which allows better Bayesian inferences in the posterior distribution.

Fig. 5 shows the autocorrelation plot within the four Markov chains for the production cycle numbers 63, 97, and 152. It is deduced that the thinning values decrease the autocorrelation efficiently.

- Divergent transitions

The Markov chain Monte Carlo estimators should converge to the true expectation values in parametric space as fast as possible, so they are reasonably accurate before exhausting finite computational resources. HMC can simulate dynamic processes throughout time via the *Störmer–Verlet* system called the "leapfrog integrator" to draw a sketch of the posterior probability surface (Hoffman and Gelman, 2014). And is important to check if there is the presence of divergent transitions during the sampling period of the model (the iterations after the burn-in). In our case, there was no divergence in the sampling after the burn-in (warm-up) period.

The step size of the leapfrog integrator used in the Hamiltonian simulation is another hyperparameter used by NUTS algorithm. HMC implementations use numerical integrators requiring a step size (discretization time interval, equivalently). A large step size implies the integrator will be unreliable and proposals for the posterior distribution will be unacceptable, a small step size implies many small steps will be taken by the integrator leading to long simulation times. Thus the algorithm needs to find a balance in step size value. Therefore, the acceptance average rate of step size during the warm-up period was 0.00979 ± 0.00235 between Markov chains for our model (Fig. 6) converging to the value found well before the 1000th iteration.

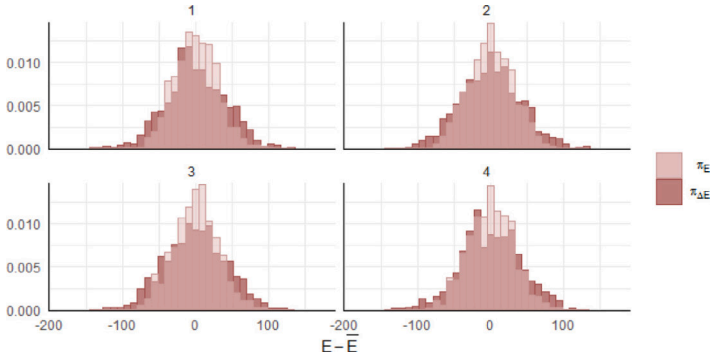


Fig. 8. NUTS Energy Diagnostic. The plot shows overlaid histograms of the (centered) marginal energy distribution π_E and the first-differenced distribution (the energy transition density) $\pi_{\Delta E}$.

- Tree-depth plot

In our warm-up period MCMC chains, the tree-depth parameter achieved the bumping up against the max value before the 1000th iteration until the hyperparameter value converged on average to 9.01 ± 0.305 between all chains (Fig. 7). This pattern is normal until the hyperparameter reaches its maximum efficient value during this tuning process.

- Bayesian fraction of missing information

The energy diagnostic plot shows overlaid histograms of the energy transition density and marginal energy distribution (Fig. 8). This means there are no discrepancies between these distributions samples, distributions are well-matched and the Hamiltonian Markov chain should perform robustly.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compag.2022.107196.

References

Agyeman, P.C., Khosravi, V., Kebonye, N.M., John, K., Boruvka, L., Vasat, R., 2022. Using spectral indices and terrain attribute datasets and their combination in the prediction of cadmium content in agricultural soil. Comput. Electron. Agric. 198, 107077.

Ali, M., Nicieza, A., Wootton, R.J., 2003. Compensatory growth in fishes: a response to growth depression. Fish. Fish. 4 (2), 147–190.

Aragón-Noriega, E.A., Mendivil-Mendoza, J.E., Alcántara-Razo, E., Valenzuela-Quiñónez, W., Félix-Ortiz, J.A., 2017. Multi-criteria approach to estimate the growth curve in the marine shrimp, Penaeus vannamei Boone, 1931 (Decapoda, Penaeidae). Crustaceana 90 (11–12), 1517–1531.

Berger, J.O., 1985. Prior information and subjective probability. In: Statistical Decision Theory and Bayesian Analysis. Springer, pp. 74–117.

Betancourt, M., 2016a. Diagnosing suboptimal cotangent disintegrations in Hamiltonian Monte Carlo. arXiv:1604.00695.

Betancourt, M., 2016b. Identifying the optimal integration time in Hamiltonian Monte Carlo. arXiv:1601.00225.

Betancourt, M., 2017. A conceptual introduction to Hamiltonian Monte Carlo. arXiv:1701.02434.

Betancourt, M.J., Girolami, M., 2013. Hamiltonian Monte Carlo for hierarchical models. arXiv:1312.0906.

Betancourt, M., Stein, L.C., 2011. The geometry of Hamiltonian Monte Carlo. arXiv:1112.4118.

Cervantes-Hernández, P., Torres-Hernández, P., Gómez-Ponce, M.A., 2017. Recruitment age of Litopenaeus vannamei (Boone, 1931)(Decapoda: Penaeidae) in the Cabeza de Toro-La Joya Buenavista Lagoon System, Oaxaca-Chiapas, México. Open J. Mar. Sci. 7 (4), 511–525.

Chávez, E., 1973. Estudio sobre la tasa de crecimiento del camarón blanco (Penaeus vannamei, Boone) de la región sur del Golfo de California. Ciencia, Mex 28 (2), 79–85.

Estrada-Pérez, A., Ruiz-Velazco, J.M., Hernández-Llamas, A., Zavala-Leal, I., Martínez-Cárdenas, L., 2016. Deterministic and stochastic models for analysis of partial harvesting strategies and improvement of intensive commercial production of whiteleg shrimp (Litopenaeus vannamei). Aquac. Eng. 70, 56–62.

Fao, 2020. The state of world fisheries and aquaculture 2020. Sustainability in action. Rome.

Fernandes, T.J., Muniz, J.A., Pereira, A.A., Muniz, F.R., Muianga, C.A., 2015. Parameterization effects in nonlinear models to describe growth curves. Acta Scientiarum. Technol. 37 (4), 397–402.

Gallardo-Collí, A., Pérez-Fuentes, M., Pérez-Rostro, C.I., Hernández-Vergara, M.P., 2020. Compensatory growth of Nile tilapia Oreochromis niloticus, L. subjected to cyclic periods of feed restriction and feeding in a biofloc system. Aquacult. Res. 51 (5), 1813–1823.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis. Chapman and Hall/CRC.

Gelman, A., Rubin, D.B., 1992. Lnference from iterative simulation using multiple sequences. Statist. Sci. 7 (4), 457–472.

Hernández-Covarrubias, V., Muñoz-Rubí, H.A., Madrid-Vera, J., Chávez-Herrera, D., 2012. Fecundidad del camarón blanco Litopenaeus vannamei de la plataforma continental de Sinaloa, Mpxico. Cienc Pesq 20 (2), 17–21.

Hernandez-Llamas, A., Ratkowsky, D.A., 2004. Growth of fishes, crustaceans and molluscs: estimation of the von Bertalanffy, Logistic, Gompertz and Richards curves and a new growth model. Mar. Ecol. Prog. Ser. 282, 237–244.

Hoffman, M.D., Gelman, A., 2014. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. J. Mach. Learn. Res. 15 (1), 1593–1623.

Holthuis, L.B., 1990. FAO species catalogue. Vol. 1 - shrimps and prawns of the world. An annotated catalogue of species of interest to Fisheries. FAO Fish. Synop. 39–46.

IBGE, 2020. Instituto Brasileiro de geografia e Estatística. Nota Técnica 48, URL https://sidra.ibge.gov.br/pesquisa/ppm/quadros/brasil/2020, Accessed: 2022-04-12.

Jaynes, E.T., 2003. Probability Theory: The Logic of Science. Cambridge University Press.

Li, X., Ata-UI-Karim, S.T., Li, Y., Yuan, F., Miao, Y., Yoichiro, K., Cheng, T., Tang, L., Tian, X., Liu, X., et al., 2022. Advances in the estimations and applications of critical nitrogen dilution curve and nitrogen nutrition index of major cereal crops. a review. Comput. Electron. Agric. 197, 106998.

Lluch, D., 1974. La pesquería de camarón de altamar en el noroeste en el noroeste: un analysis biologica pesquera. Serie Informativa INP S 1, 16.

Lopez, S., France, J., Gerrits, W., Dhanoa, M., Humphries, D., Dijkstra, J., 2000. A generalized Michaelis-Menten equation for the analysis of growth. J. Anim. Sci. 78 (7), 1816–1828.

Luo, X., Kareem, A., 2020. Bayesian deep learning with hierarchical prior: Predictions from limited and noisy data. Struct. Saf. 84, 101918.

Mauritzen, J., 2020. Are solar panels commodities? A Bayesian hierarchical approach to detecting quality differences and asymmetric information. European J. Oper. Res. 280 (1), 365–382.

Metropolis, N., Ulam, S., 1949. The monte carlo method. J. Amer. Statist. Assoc. 44 (247), 335–341.

Michaelis, L., Menten, M., 1913. Die Kinetik der Invertinwirkung Biochemische Zeitschrift. Biochemische Zeitschrift.

Moala, F., Penha, D., 2016. Elicitation methods for Beta prior distribution. Revista Brasileira de Biometria (ISSN: 1983-0823) 34 (1), 49–62, URL http://www.biometria.ufla.br/index.php/BBJ/article/view/91.

Mohanty, R., 2015. Effects of feed restriction on compensatory growth performance of Indian major carps in a carp–prawn polyculture system: a response to growth depression. Aquacult. Nutr. 21 (4), 464–473.

Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective. MIT Press.

Neal, R.M., 1993. Probabilistic Inference using Markov Chain Monte Carlo Methods. Department of Computer Science, University of Toronto Toronto, ON, Canada.

Neal, R.M., 2012. Mcmc using Hamiltonian dynamics. http://dx.doi.org/10.1201/b10905, arXiv:1206.1901.

Park, S.Y., Bera, A.K., 2009. Maximum entropy autoregressive conditional heteroskedasticity model. J. Econometrics 150 (2), 219–230.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Ramos-Cruz, S., 2011. Relaciones biométricas del camarón blanco Litopenaeus vannamei Boone 1931 (Decapoda, Penaeidae), para la región del Golfo de Tehuantepec, México. CICIMAR Oceánides 26 (2), 71–75.

Ruiz-Velazco, J.M., Hernández-Llamas, A., Gomez-Muñoz, V.M., 2010. Management of stocking density, pond size, starting time of aeration, and duration of cultivation for intensive commercial production of shrimp Litopenaeus vannamei. Aquac. Eng. 43 (3), 114–119.

Salles, T., Beijo, L.A., Nogueira, D., Almeida, G.C., Martins, T.B., Gomes, V.S., 2020. Modelling the growth curve of Santa Ines sheep using Bayesian approach. Livestock Sci. 104115.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423.

Shi, D., Tong, X., 2020. Mitigating selection bias: a Bayesian approach to two-stage causal modeling with instrumental variables for nonnormal missing data. Sociol Methods Res. 0049124120914920.

Singh, V.P., Rajagopal, A., Singh, K., 1986. Derivation of some frequency distributions using the principle of maximum entropy (POME). Adv. Water Resour. 9 (2), 91–106.

Stan Development Team, 2020. Rstan: the R interface to Stan. URL http://mc-stan.org/, R package version 2.21.2.

Tagliafico, A., Rangel, S., Kelaher, B., Christidis, L., 2018. Optimizing heterotrophic feeding rates of three commercially important scleractinian corals. Aquaculture 483, 96–101.

Tian, X., Dong, S., 2006. The effects of thermal amplitude on the growth of Chinese shrimp Fenneropenaeus chinensis (Osbeck, 1765). Aquaculture 251 (2–4), 516–524.

Tian, X., Leung, P., Hochman, E., 1993. Shrimp growth functions and their economic implications. Aquac. Eng. 12 (2), 81–96.

VanDerwerken, D., Schmidler, S.C., 2017. Monitoring joint convergence of MCMC samplers. J. Comput. Graph. Statist. 26 (3), 558–568.

Wasielesky Jr., W., Froes, C., Fóes, G., Krummenauer, D., Lara, G., Poersch, L., 2013. Nursery of litopenaeus vannamei reared in a biofloc system: the effect of stocking densities and compensatory growth. J. Shellfish Res. 32 (3), 799–806.

Yu, R., Leung, P., 2010. A Bayesian hierarchical model for modeling white shrimp (Litopenaeus vannamei) growth in a commercial shrimp farm. Aquaculture 306 (1–4), 205–210.

Yu, R., Leung, P., Bienfang, P., 2006. Predicting shrimp growth: artificial neural network versus nonlinear regression models. Aquac. Eng. 34 (1), 26–32.

Zhou, S., Martin, S., Fu, D., Sharma, R., 2020. A Bayesian hierarchical approach to estimate growth parameters from length data of narrow spread. ICES J. Mar. Sci. 77 (2), 613–623.

ARTIGO 2 - ***Modeling the growth of Pacific white shrimp (Litopenaeus vannamei) using the new Bayesian hierarchical approach based on correcting bias caused by incomplete or limited data***

Redigido conforme as normas da revista *Ecological Informatics* (versão em processo de editoração).

Modeling the growth of Pacific white shrimp (*Litopenaeus vannamei*) using the new Bayesian hierarchical approach based on correcting bias caused by incomplete or limited data

Carlos Antônio Zarzar[a,b,*], Tales Jesus Fernandes[b], Izabela Regina Cardoso de Oliveira[b]

[a] *Universidade Federal do Oeste do Pará (UFOPA), Av. Maj. Francisco Mariano CEP: 68220-000, Cidade Alta, Monte Alegre, Pará, Brasil.*
[b] *Universidade Federal de Lavras (UFLA), 3037 CEP: 37200-900, Lavras, Minas Gerais, Brasil*

Abstract

The revenue, costs, and profit of an aquaculture farm are based on the weight of animal protein sold. Thus, there is a relationship between the economic and zootechnical indexes to the growth model of animals in a production system. The growth modeling of cultivated organisms can be used as a production management tool, allowing estimates of anticipated size at harvest, waste outputs, as well as nutrient and feed requirements, helping in decision-making in the face of Aquaculture 4.0. It is important to emphasize that recent research has indicated the possible underestimation of parameters in nonlinear growth models due to the characteristic of incomplete or limited data in aquaculture. Therefore, the objective of this research was to model the growth of Pacific white shrimp (*Litopenaeus vannamei*) in an industrial-scale shrimp farm in northeastern Brazil. Based on the Bayesian methodology for correcting this bias, six nonlinear hierarchical growth models were evaluated in this research (Morgan-Mercer-Flodin, Michaelis-Menten, Weibull, von Bertalanffy, Gompertz, and logistic growth equation) and fitted to real data from a shrimp farm in northeastern Brazil. The model was validated based on the predictive capacity (accuracy) to forecasting shrimp growth at different hierarchical levels. Finally, for one of the main expected results, a sensitivity analysis was performed to compare different treatments according to the new approach. The Weibull growth equation stands out as the best among all those studied (WAIC= 2661.3, LOO-IC= 2705.2). Although it presented a poor fit for the hierarchical population level, good predictions were realized at the pond level and at the production cycle level. The dataset was split into fit and homologation subsets for a model validation analysis which showed an accuracy of 95.76% and 85.71% at the pond

*Corresponding author
Email address: carolos.zarzar@ufopa.edu.br (Carlos Antônio Zarzar)

and production cycle levels respectively. The proposed method detected subtle differences between production cycles, which would be imperceptible if analyzed using the zootechnical indices usually practiced in shrimp farms. The new approach can contribute to improving products, processes, and decision-making in aquaculture management.

Keywords:
Crustaceans; Mixed model; Multilevel model; Bayesian model; Hamiltonian Monte Carlo; Aquaculture.

---

1. Introduction

Statistical and mathematical modeling has been used promising in confined animal production, especially in shrimp farming, providing an understanding of cultivation-scale metabolic processes (de Melo Filho et al., 2020), growth performance (Yu et al., 2006; Yu and Leung, 2010; Zarzar et al., 2022), partial harvesting strategies (Estrada-Pérez et al., 2016; Ruiz-Velazco et al., 2021), stocking density, pond size, and time aeration (Ruiz-Velazco et al., 2010), aiding problem-solving in real-world applications. Among several models to describe growth, noteworthy are the nonlinear models due to the biological interpretation of their parameters (e.g. maximum relative growth rate, first maturation size, maximum asymptotic size, inflection point), and its parsimoniousness (i.e., usually fewer parameters involved) (Bates and Watts, 2007), which are interesting characteristics for farming.

Animals growth modeling in aquaculture can be used as a production management tool allowing anticipated size at harvest, estimates of waste outputs as well as nutrient and feed requirements (Dumas et al., 2008; Einen et al., 1995; Bailey and Alanärä, 2001; Cho and Bureau, 1998). It has allowed quantifying the influences of certain factors on growth, helped harvest-related management decisions, and has been used to calculate productivity and economic performance (Cacho et al., 1991; Yu et al., 2006; Martinez and Seijo, 2001). Given its importance and its wide application, mainly in shrimp farming, growth models in conjunction with advanced information technology (IT) systems and associated emerging technologies (Pache et al., 2022; Abinaya et al., 2021) can become essential tools for decision making, reducing costs, enabling growth, and expansion of production in a competitive market (Sanchez and Gonzalez, 2021).

Among the nonlinear models, the most widely used to represent the animals' growth are the models classified as sigmoids (the S-shaped growth curve), whose parameters have a clear meaning, as well their units are associated with their biological definition. Polynomial models commonly bring the disadvantages of using more parameters that are difficult to interpret (Pinheiro and Bates, 2006) and usually their predictions tend to be less robust in contrast to sigmoidal nonlinear models, especially outside the range of the observed data domain. However, sigmoid models' disadvantages when compared to linear models are that they can be less flexible (Miguez et al., 2018) and generally there is no analytical solution for estimating the parameters.

Therefore, the choice of the model is very important. In the literature, there is a satisfactory number of sigmoidal models that fit well to the growth phenomenon. Among them, it can be cited the Logistic growth equation, Gompertz, von Bertalanffy, Richard, Weibull, Morgan-Mercer-Flodin, and including Michaelis-Menten growth equation (Tian et al., 1993; Yu et al., 2006; Yu and Leung, 2010; Ruiz-Velazco et al., 2010; Estrada-Pérez et al., 2016; Aragón-Noriega et al., 2017). In the realm of statistical modeling selection, it is not always necessary to limit ourselves to a single selected model that is deemed the most suitable for a given dataset. Instead, sometimes it is viable to select a set of models (those with near selection indices) that are deemed appropriate to represent the phenomenon under study. However, it is typically advantageous to prioritize the selection of a model based on its track record of successful application in similar contexts, the property of the biological meanings of the parameters, and, most importantly, the estimated predictive accuracy as the criteria to evaluate, understand, and select models. (Gelman et al., 2014).

Recently, hypotheses have been raised that traditional methods of modeling using data classified as incomplete or limited, could lead to underestimations of nonlinear growth model parameters. An alternative method, based on the Bayesian approach, was proposed to correct this problem (Zarzar et al., 2022). An essential aspect of Bayesian theory involves the selection of the prior density function. However, when researchers encounter challenges in eliciting a priori probability densities and expert opinions on model parameters are not evident, a highly recommended approach for comprehending the implications of a prior distribution within a generative model is Prior Predictive Checking (PPC) (Box, 1980; Gabry et al., 2019; Gelman et al., 2020).

The prior predictive check feature is based on simulated data from the proposed priors distribution. Specifying proper priors for all parameters in the model, it is expected that the generative Bayesian model yields a joint prior distribution on parameters and simulated data, and consequently a prior marginal distribution for the simulated data. Thus, we could visualize simulations from the prior marginal distribution of the simulated data to assess the consistency of the chosen priors base on domain knowledge (Gabry et al., 2019). In this way, the observed data information is preserved, once fake data can be almost as valuable as real data for building the model (Gabry et al., 2019). All this provides a way to refine the model without using the data multiple times (Gelman et al., 2020).

Therefore, based on the workflow Bayesian methodology (Chiu et al., 2017; Gabry et al., 2019; Gelman et al., 2020) we aim to compare different sigmoid models for modeling the Pacific white shrimp growth (*Litopeaneus vannamei*) fitted to the real data from commercial farms in northeastern Brazil, using the proposed bias correction method by Zarzar et al. (2022). Furthermore, we investigate the method's sensitivity in detecting differences between treatments through projected growth curves and estimates of model parameters.

2. Materials and Methods

2.1. Dataset

The data used in this study was obtained from a commercial shrimp farm located in Northeast Brazil. The farm, which focused on Pacific white shrimp (*Litopenaeus vannamei*), operated a total of 55 ponds ranging in size from 1.5 to 19.3 hectares during the years 2017 and 2018. For this research, weekly sampling data were collected from 40 specific grow-out ponds in 205 production cycles within the specified timeframe. The duration of these production cycles varied between 5 weeks (38 days) and 15 weeks (102 days). The initial stocking density (juvenile shrimps weighing between 0.2g) ranged from 2 to 30 individuals per square meter. The shrimps were then harvested when they reached an average weight of 8.43g with a standard deviation of 0.72g. In total, the dataset comprises 1820 observations.

2.2. Bayesian workflow

Any methodology based on the Bayesian perspective should follow good practices of Bayesian statistics. This article was based on the workflow Bayesian methodology developed by Chiu et al. (2017); Gabry et al. (2019); Gelman et al. (2020), over the last few years. The Bayesian workflow is concerned with three critical points of Bayesian statistics: definition of the model, inference, and verification/model improvement. These critical points involve several moments in Bayesian modeling, either before, during, or after fitting a model. The workflow Bayesian methodology was split into sections and described below in order to favor the understanding of the research.

2.3. Before fitting a model

2.3.1. Candidate nonlinear models

The average weight (grams) of the Pacific white shrimp (*Litopenaeus vannamei*) as a function of time was fitted to six nonlinear sigmoid growth curves. Table 1 shows all the nonlinear equations investigated in this work and their respective parameterizations, as well as the parameter restrictions of each one. The Bayesian hierarchical model approach using the Hamiltonian Monte Carlo method (HMC) from Markov Chain Monte Carlo (MCMC) to solve the underestimation parameter problem derived from incomplete or limited data, was adopted. We followed the same methodologies, including the numerical computational method and model assumptions pointed out by Zarzar et al. (2022).

Table 1: The nonlinear equations $f(t, \theta_n)$ (sigmoidal growth curves) investigated for growth of the Pacific white shrimp (*Litopenaeus vannamei*) in farming settings. The parameterization and parameter restrictions for growth were defined in the first moment. t is the time, θ_n is generalized parameter vector, where n is the index number of nonlinear parameter with $n = 1, 2, 3$ some case 4

Function name	**Mathematical expression** ($f(t, \theta_n)$)	**Parameters restrictions** (θ_n)
Morgan-Mercer-Flodin (MMF)	$f(t,\theta_n) = \alpha - \frac{\alpha - w_0}{1 + (\kappa \cdot 10^{-4} \cdot\ t)^{\delta}}$	$\theta_1 = \alpha > 0;\ \theta_2 = \kappa > 0;$ $\theta_3 = \delta \in \mathbb{R}_+^* \mid \delta \neq 1$ and $\theta_4 = w_0 \geq 0 \mid w_0 = 0.2$
Michaelis-Menten Generalized	$f(t,\theta_n) = \frac{w_0\ \beta^{\kappa} + \alpha\ t^{\kappa}}{\beta^{\kappa} + t^{\kappa}}$	$\theta_1 = \alpha > 0;\ \theta_2 = \kappa > 0;$ $\theta_3 = \beta > 0$ and $\theta_4 = w_0 \geq 0 \mid w_0 = 0.2$
Weibull growth	$f(t,\theta_n) = \alpha\ (1 - exp(-\beta \cdot 10^{-4} \cdot\ t^{\kappa})) + w_0$	$\theta_1 = \alpha > 0;\ \theta_2 = \kappa > 1;$ $\theta_3 = \beta > 0$ and $\theta_4 = w_0 \geq 0 \mid w_0 = 0.2$
von Bertalanffy	$f(t,\theta_n) = \alpha\ (1 - exp(-\kappa \cdot 10^{-4} \cdot\ (t + \beta)))^3$	$\theta_1 = \alpha > 0;\ \theta_2 = \kappa > 0$ and $\theta_3 = \beta \in \mathbb{R}$
Gompertz function	$f(t,\theta_n) = \alpha\ exp(-exp(\kappa\ (\beta - t)))$	$\theta_1 = \alpha > 0;\ \theta_2 = \kappa > 0$ and $\theta_3 = \beta \in \mathbb{R}$
Logistic function	$f(t,\theta_n) = \frac{\alpha}{1 + exp(\kappa\ (\beta - t))}$	$\theta_1 = \alpha > 0;\ \theta_2 = \kappa > 0$ and $\theta_3 = \beta \in \mathbb{R}$

2.3.2. Prior Predictive checks (PPC_{Prior})

The PPC_{Prior} is based on data simulation which is generated from the interaction between proposed priors distributions and the likelihood. This process is called the prior predictive distribution and shows how the model behaves before using the data. Furthermore, generating prior predictive distributions repeatedly helps us to check whether the priors make sense.

The PPC_{Prior} was performed from the model assumption (Section 2.4) for each growth curve described in Table 1 and the proposed priors distributions (Table 2). 1,000 random and independent samples were generated from the proposed prior distributions for each parameter of the nonlinear model analyzed. Then the posterior distributions were calculated from the joint priors. Through the graphical visualizations of this proposed generative Bayesian model, the consistency of the choice of the proposed priors was evaluated with respect to the known domain of the dependent variable under study (shrimp weight grams over growing days). Joint priors allow us to control the overall complexity of larger parameter sets, which helps generate more sensible prior predictions that would be hard or impossible to achieve with independent priors (see, e.g., Piironen et al. (2017), and Zhang et al. (2020)).

Formally, to calculate the prior predictive distribution $p(w_{gener})$, we generated data w_{gener} over proposed priors distributions $p(\theta)$, with such parameter values ($\theta \in \Theta$) and likelihood $p(w_{gener} \mid \theta)$, which is written as follows:

$$\begin{aligned} p(w_{gener}) &= \int_{\theta \in \Theta} p(w_{gener}, \theta)\, d\theta \\ &= \int_{\theta \in \Theta} p(w_{gener} \mid \theta)\, p(\theta)\, d\theta \end{aligned}$$

Table 2: The proposed priors distributions in Bayesian workflow for each parameter of the nonlinear sigmoid growth curves $f(t, \theta_{n|ijk})$ for the hierarchical model in modeling the average shrimp weight (grams) of the Pacific white shrimp (*Litopenaeus vannamei*) as a function of time (days)

Function name	**Parameter**	**Group Level**	**Prior distribution**
Morgan-Mercer-Flodin (MMF)	$\theta_{1\|i} = \alpha_i$	Population	Normal$(85, 5)$
	$\theta_{2\|ijk} = \kappa_{ijk}$	Population	Normal$(60, 12)$
	$\theta_{3\|ijk} = \delta_{ijk}$	Population	Skew Normal$(1.89, 0.15, -2.0)$
	$\tau_{\theta_{2\|jk}} = \tau_{\kappa_{jk}}$	Pond	Normal$(0, 7.5)$
	$\tau_{\theta_{2\|k}} = \tau_{\kappa_k}$	Cycle-Pond	Normal$(0, 7.5)$
	$\tau_{\theta_{3\|jk}} = \tau_{\delta_{jk}}$	Pond	Normal$(0, 0.1)$
	$\tau_{\theta_{3\|k}} = \tau_{\delta_k}$	Cycle-Pond	Normal$(0, 0.1)$
Michaelis-Menten Generalized	$\theta_{1\|i} = \alpha_i$	Population	Normal$(85, 1)$
	$\theta_{2\|ijk} = \kappa_{ijk}$	Population	Skew Normal$(2.6, 0.15, -2, 12)$
	$\theta_{3\|ijk} = \beta_{ijk}$	Population	Normal$(200, 20)$
	$\tau_{\theta_{2\|jk}} = \tau_{\kappa_{jk}}$	Pond	Normal$(0, 0.015)$
	$\tau_{\theta_{2\|k}} = \tau_{\kappa_k}$	Cycle-Pond	Normal$(0.08, 0.2)$
	$\tau_{\theta_{3\|jk}} = \tau_{\beta_{jk}}$	Pond	Normal$(4, 3)$
	$\tau_{\theta_{3\|k}} = \tau_{\beta_k}$	Cycle-Pond	Normal$(30, 3)$
Weibull growth	$\theta_{1\|i} = \alpha_i$	Population	Normal$(85, 5)$
	$\theta_{2\|ijk} = \kappa_{ijk}$	Population	Normal$(1.5, 0.05)$
	$\theta_{3\|ijk} = \beta_{ijk}$	Population	Skew Normal$(1.3, 5, 6)$
	$\tau_{\theta_{2\|jk}} = \tau_{\kappa_{jk}}$	Pond	Normal$(0, 0.008)$
	$\tau_{\theta_{2\|k}} = \tau_{\kappa_k}$	Cycle-Pond	Normal$(0, 0.005)$
	$\tau_{\theta_{3\|jk}} = \tau_{\beta_{jk}}$	Pond	Normal$(0, 0.0005)$
	$\tau_{\theta_{3\|k}} = \tau_{\beta_k}$	Cycle-Pond	Normal$(0, 0.0005)$
von Bertalanffy	$\theta_{1\|i} = \alpha_i$	Population	Normal$(85, 1)$
	$\theta_{2\|ijk} = \kappa_{ijk}$	Population	Normal$(1.2, 0.3)$
	$\theta_{3\|ijk} = \beta_{ijk}$	Population	Normal$(6, 0.5)$
	$\tau_{\theta_{2\|jk}} = \tau_{\kappa_{jk}}$	Pond	Normal$(0, 0.06)$
	$\tau_{\theta_{2\|k}} = \tau_{\kappa_k}$	Cycle-Pond	Normal$(0, 0.06)$
	$\tau_{\theta_{3\|jk}} = \tau_{\beta_{jk}}$	Pond	Normal$(0, 0.5)$
	$\tau_{\theta_{3\|k}} = \tau_{\beta_k}$	Cycle-Pond	Normal$(0, 0.5)$
Gompertz function	$\theta_{1\|i} = \alpha_i$	Population	Normal$(85, 1)$
	$\theta_{2\|ijk} = \kappa_{ijk}$	Population	Normal$(0.020, 0.003)$
	$\theta_{3\|ijk} = \beta_{ijk}$	Population	Normal$(100, 5)$
	$\tau_{\theta_{2\|jk}} = \tau_{\kappa_{jk}}$	Pond	Normal$(0, 0.0005)$
	$\tau_{\theta_{2\|k}} = \tau_{\kappa_k}$	Cycle-Pond	Normal$(0, 0.0005)$
	$\tau_{\theta_{3\|jk}} = \tau_{\beta_{jk}}$	Pond	Normal$(0, 5)$
	$\tau_{\theta_{3\|k}} = \tau_{\beta_k}$	Cycle-Pond	Normal$(0, 8)$
Logistic function	$\theta_{1\|i} = \alpha_i$	Population	Normal$(85, 1)$
	$\theta_{2\|ijk} = \kappa_{ijk}$	Population	Normal$(0.063, 0.001)$
	$\theta_{3\|ijk} = \beta_{ijk}$	Population	Normal$(100, 10)$
	$\tau_{\theta_{2\|jk}} = \tau_{\kappa_{jk}}$	Pond	Normal$(0.001, 0.0001)$
	$\tau_{\theta_{2\|k}} = \tau_{\kappa_k}$	Cycle-Pond	Normal$(0.002, 0.001)$
	$\tau_{\theta_{3\|jk}} = \tau_{\beta_{jk}}$	Pond	Normal$(5, 5)$
	$\tau_{\theta_{3\|k}} = \tau_{\beta_k}$	Cycle-Pond	Normal$(10, 8)$
Family-specific Distribution	σ_i	Population	Student's t$(3, 0, 3.9)$

2.4. Fitting a model

2.4.1. The Bayesian hierarchical model

We used the Bayesian hierarchical model approach with the Hamiltonian Monte Carlo numeric method (HMC) to solve the underestimation parameter problem, derived from incomplete or limited data. Following the same methodologies, including the numerical computational method and model assumptions carried out by Zarzar et al. (2022).

In general terms, we can think about the model as a generating-data mechanism for our observations on the weight of shrimp (w) over time (t) through a stochastic process. In this way, a probability distribution of the data (D) is assumed and the parameters by which the process is governed. Among the parameters, they distinguish between parameters that are direct functions of variables in the data (Θ) and family-specific probability distribution (Φ).

$$w \sim D(f(t, \Theta), \Phi)$$

For this particular research, in the context of the multilevel model, we assumed that the generating-data mechanism follows a normal distribution with mean μ and constant standard deviation σ:

$$w_{ijk} \sim Normal(\mu_{ijk}, \sigma_i) \tag{1}$$

where i is the index for the ith observation from population level, j is the index for the jth tank (pond) group from hierarchy level, k is the index for the kth cycle production group from hierarchy level nested with pond group, where $\mu_{ijk} = f(t, \theta_{n|ijk})$ being $f(t, \theta_{n|ijk})$ a nonlinear (sigmoid) function of growth (Table 1), θ_n is the vector of nonlinear parameters, n is the index for the nth parameter of growth sigmoid function. Thus, expression 1 can be written in the following hierarchical form:

Population (level one):

$$w_{ijk} \sim Normal(f(t, \theta_{n|ijk}), \sigma_i)$$

Pond group (level two):

$$\theta_{n|ijk} \sim Normal(\theta_{n|jk}, \tau_{\theta_{n|jk}})$$

Cycle-Pond group (level three):

$$\theta_{n|jk} \sim Normal(\theta_{n|k}, \tau_{\theta_{n|k}})$$

where the τ_{θ_n} parameter represents the standard deviation of the θ_n parameter for each hierarchy.

It is important to point out that the proposed Bayesian Hierarchical model

thoughtful the parameter $\theta_{1|i}$ equals α_i, which represents the theoretical asymptotic weight parameterized in all nonlinear sigmoid functions considered in this research, only at the population level. The weight at the limit of time tending to zero was taken on as a constant equal to 0.2 ($\theta_{4|i} = w_0 = \lim_{t\to 0} f(t, \theta_{n|ijk}) = 0.2$) for the functions that have this parameter.

2.4.2. Computational numerical integration method

The model was fitted using the Stan probabilistic programming language (Stan Development Team, 2020) with R interface, which implements the Hamiltonian Monte Carlo algorithm (HMC) [it belonging to the Markov chain Monte Carlo (MCMC) family] (Neal, 2011), particularly the No-U-Turn Sampler (NUTS) which is a variation of the HMC algorithm (Hoffman and Gelman, 2014). Furthermore, all analysis was performed in the free R program (R Core Team, 2021).

The models were run with 4 Markov chains with 5,000 iterations per chain. The warmup period (aka burning) was configured with 2,500 iterations per chain (50% on iteration number). The period for saving samples (aka thinning in MCMC) was set with 1 lag (thinning rate). Thus, we obtained 2,500 iterations per chain post-warmup period, which corresponds to 10,000 total interactions post-warmup draws.

2.4.3. Health Diagnosis and quality of the MCMC chains convergences

At the end of all iterations of the MCMC method, it is expected that the 4 chains have reached the desired convergence. This condition was verified through graphical analysis, and the Gelman–Rubin potential scale reduction factor ($\hat{R}$), in which it must be around 1 indicating the Markov chains have converged, but it must not exceed 1.01 nor less than 0.99 (see Kruschke (2014) and Gelman and Rubin (1992) for details). $\hat{R}$ is the statistic that monitors the convergence of the chains to the equilibrium of the target distribution through the ratio of the average variance of samples within each chain to the variance of the pooled samples across chains. If all chains are at equilibrium, these will be the same, and $\hat{R} = 1$. If the chains have not converged to a common distribution, $\hat{R} \neq 1$.

We also checked the effective number of independent samples n_{eff}, indicating lower autocorrelation and higher efficiency of the convergence (McElreath, 2020). If there is autocorrelation, the effective sample size will be smaller than the total number of iterations (N). In this research, $n_{eff}/N = 0.10$ was used as a reference and the estimated model parameters with values below it may indicate that the MCMC sampling after warmup may have some autocorrelation.

2.5. After fitting the model

2.5.1. Comparison and models selection

After fitting a Bayesian hierarchical model for each growth curve, the comparison between models was based on predictive accuracy athwart the estimation of Expected log Pointwise Predictive Density (ELPD), computed through leave-one-out cross-validation (LOO-CV). It is the sum of the log-probability scores

for the given data except for one data point left out at a time (Vehtari et al., 2017; McElreath, 2020). In addition, the Information Criterion of Leave-one-out cross-validation (LOO-CI) value, obtained through the Pareto-smoothed importance sampling (PSIS) approximation method, and The Watanabe-Akaike Information Criterion (WAIC) (Watanabe and Opper, 2010) were used in order to select a set of models (or one model) considered suitable for the phenomenon under study.

2.5.2. Diagnostic and assessment of the selected model

For the models' evaluation and diagnostic, we also followed the Bayesian workflow suggested by Gelman et al. (2020) who recommends using constructed simulated data to find and understand problems based on predictive performance [*posterior predictive checks* $PPC_{Posterior}$ e.g., (Gelman et al., 2013)]. Fundamentally this analysis is made through graphical visualizations that allow us to identify existing problems and understand how such problems affect Bayesian inference. We first fit the model to the data to get the posterior distribution of the latent parameters. The fitted model induces a distribution of future data conditioned on the observations generating synthetic datasets. Then the $PPC_{Posterior}$ consists of the check whether the simulated datasets are close to the observed dataset when summarized through a statistic of interest. Denote the distribution of this replication by $p(\tilde{w} \mid f(t,\theta), w)$:

$$p(\tilde{w} \mid f(t,\theta), w) = \int_{\theta \in \Theta} p(\tilde{w} \mid f(t,\theta), \theta)\, p(\theta \mid f(t,\theta), w)\, d\theta$$

where the notation w is the observed data, $f(t,\theta)$ for the assumed nonlinear model, t time, θ unknown model parameters, $\tilde{w}$ the replicated data that can be observed.

Another method used for evaluating the model fit and its predictive potential to new data was the difference between the theoretically expected log pointwise predictive density for a new dataset (ELPD) and the non-cross-validated log posterior predictive density. Under certain regularity conditions, it can be asymptotically interpreted as the Effective Number of Parameters. In short, it describes the complexity of the model, that is, how much more difficult it is to predict future data than the observed data for the proposed model (Vehtari et al., 2017). For a well-fitted model, the effective number of parameters should be less than the sample size and also less than the total number of parameters, featuring expected behavior. Otherwise, the model has a very weak predictive capability pointing to possible severe model misspecification.

All these indexes for comparison, diagnostic, assessment, and model selection were presented in the Table 4 to facilitate the analysis.

2.6. Validating the selected model

In order to validate the model from the prediction of new data (data not used in the modeling), 20% (41 production cycles) of the data were separated for validation of the model. Due to the hierarchical structure of the data,

the predictive capacity of the multilevel model under evaluation may take into account different hierarchical levels depending on the main interest. Considering the productive scenario of shrimp farms, the main interest is to predict the slaughter weight moments before harvesting (around 21 days before) or predict the animal's growth even before populating the pond. Therefore, the validation data (41 production cycles) were split into two validation subsets, one to assess the predictive capacity of the model at the tank level, and another to assess the predictive capacity at the production cycle level. Thus, 21 production cycles were separated to present the model's accuracy in predicting 3 biometrics before harvesting, and 20 production cycles to present the predictive capability of new production cycles. Predictions for these conditions and observations were performed, and estimates of 95% credibility intervals from the posterior of the model were calculated and compared with the true values. The results were presented by graphs and estimated values of the model accuracy in Table 6.

2.7. Model sensitivity in detecting treatment differences

To measure the sensitivity of the method in identifying subtle differences in growth between production cycles, cultures that would normally be classified as similar based on traditional zootechnical indices used on the farm were compared. The cultures were grouped by their similarity in growth, for example, group 1: the production cycles that reached 8.2 grams within 56 to 58 days of cultivation; group 2: with 7.5 grams in 55 to 57 days of cultivation, among other groups described in Table 3. This table also brings similarities with the zootechnical index of average weekly weight gain and Specific Growth Rate (SGR).

Table 3: Production cycle grouped with similarity zootechnical performance of Pacific white shrimp (*Litopenaeus vannamei*) in a shrimp farm in northeastern Brazil. Variables cultivation days, tank size, initial stocking density, Feed Conversion Ration (FCR), and survival are presented in the respective order of each production cycle

Group Number	Days	Weight[1]	Cycle	Pond	Tank Size[2]	Initial Stocking Density[3]	Average weekly weight gain	SGR[4]	Survival (%)	FCR[5]
1	56 - 58	8.2	204 - 47	28 - 19	2.9-5	10.3 - 5	1.07 - 1.00	6.63 - 6.40	39 - 35	0.72 - 0.61
2	55 – 56 56 – 57	7.5	180 – 61 172 – 80	19 – 33 22 – 21	5 – 3 1.5 – 2.5	6 – 6 9.3 – 8	0.89 – 0.93 0.92 – 0.89	6.59 – 6.47 6.47 – 6.36	47 – 70 50 – 37	0.62 – 0.91 0.64 – 0.86
3	53 – 56	7.3	111 – 170	34 – 17	2 – 2	10 – 9	0.95 - 0.92	6.78 - 6.43	53 – 27	0.57 – 1.17
4	84 – 85	9.4	201 – 196	22 – 17	1.5 – 2	12 – 7.5	0.76 - 0.75	4.58 - 4.53	76 – 75	0.81 – 1.72
5	62 – 63	7	105 – 171	21 – 23	2.5 – 1.8	10 – 8.9	0.82 - 0.73	5.73 - 5.64	57 – 26	0.51 – 1.22
6	79 – 79	7	11 – 55	11 – 27	5 – 2	5 – 4	0.59 - 0.59	4.50 - 4.50	70 – 88	0.96 – NA[6]
7	50 – 52 53	7.8	119 – 52 36	2 – 24 3	5 – 3 9.4	6 – 6.7 8.9	1.06 – 0.98 0.98	7.33 – 7.04 6.91	54 – 47 24	0.45 – 0.93 0.87
8	56 – 57 57 – 57	8.5	203 – 77 145 – 5	37 – 10 2.95 – 5	2.6 – 4.5 12.9 – 8.1	10.8 – 10 2.9 – 5	1.02 – 0.99 0.98 – 1.00	6.69 – 6.58 6.58 – 6.58	32 – 40 52 – 74	0.56 – 0. 61 NA – 0.39
9	58 – 59	7.6	109 – 202	33 – 32	3 – 2.7	11.7 – 10	0.88 - 0.86	6.27 - 6.16	25 – 28	1.59 – 1.05

[1] Animal weight at the end of cultivation (harvest weight)
[2] Size of the tank (pond) in hectare
[3] individuals number / m^2
[4] SGR - Specific Growth Rate (% $\cdot d^{-1}$) = $\frac{ln(w_t) - ln(w_0)}{t} 100$
[5] FCR - Feed Conversion Ratio
[6] NA - Not Available

3. Results

3.1. Before fitting a model

3.1.1. Prior predictive checking

Figure 1 shows the prior predictive check for the Morgan-Mercer-Flodin, Michaelis-Menten, Weibull growth equation, von Bertalanffy, Gompertz, and the Logistic growth function based on the prior probability densities proposed in Table 2. The results indicate that the priors elicited were plausible due to the draws from the prior marginal distribution for the generated data which are within the domain knowledge. The results also cover any random future dataset that could plausibly be expected for the phenomenon. To illustrate these results, a vertical red line was placed on the graphics (Figure 1). It represents the known domain expected for the Pacific white shrimp (*Litopenaeus vannamei*) with a weight of around 49 days of cultivation. In an extensive cultivation system in the tropical region, the shrimp weight is expected to range from 0 to 20 grams until this cultivation day. It can be noted that the prior predictive distribution generated data on the prior proposals beyond the expected boundaries giving a safe margin on the uncertainty of shrimp growth.

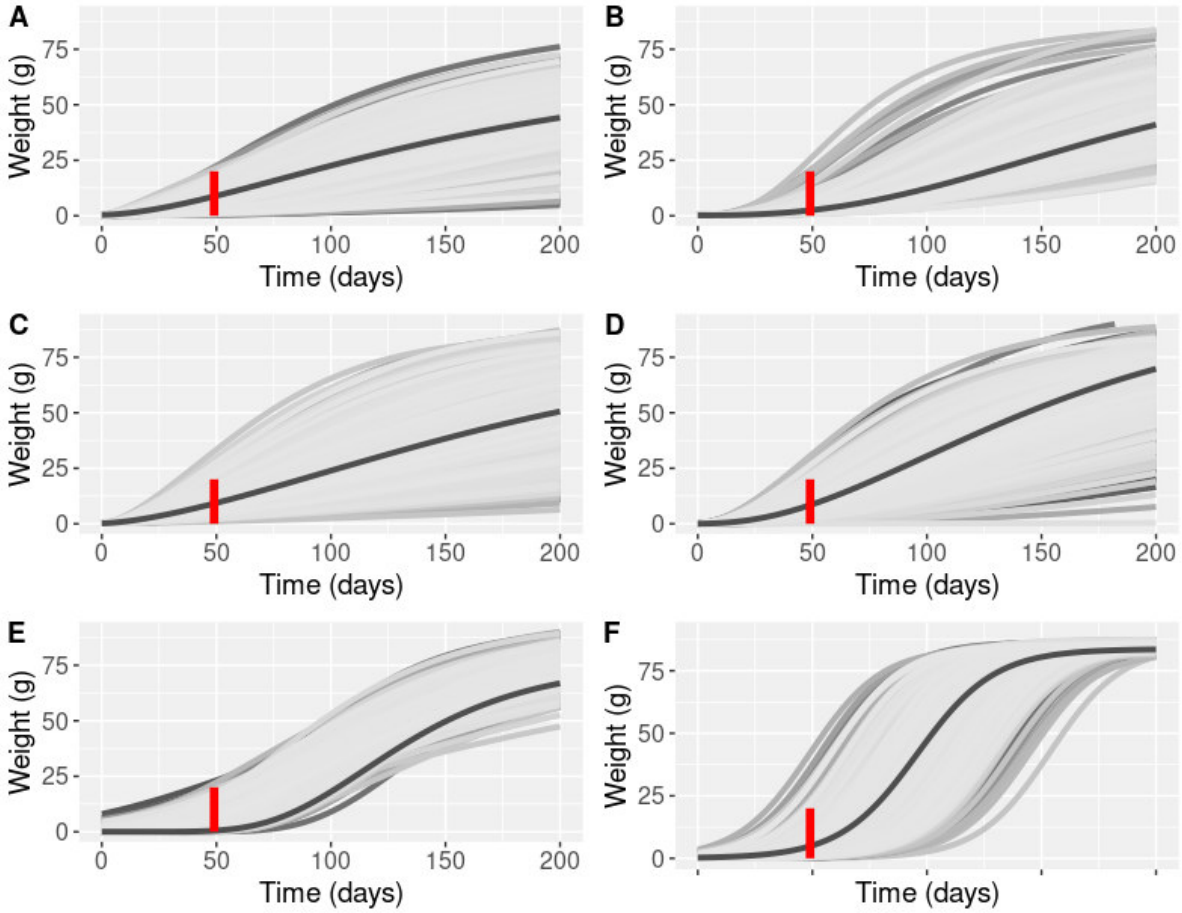


Figure 1: Prior predictive check (PPC_{prior}) feature in workflow Bayesian analysis for A) Morgan-Mercer-Flodin (MMF), B) Michaelis-Menten, C) Weibull growth equation, D) von Bertalanffy, E) Gompertz and F) the Logistic growth function based on the prior probability densities proposed in Table 2. The red vertical line is the known domain expected of the Pacific white shrimp (*Litopenaeus vannamei*) around 49 days of cultivation in an extensive cultivation context

3.2. Fitting a model

3.2.1. Health Diagnosis and quality of the MCMC chains convergences

In general, there was no evidence of non-convergence in the chains sampled by the MCMC (HMC) method. The potential scaling factor for all models showed a healthy convergence ($0.9996 < \hat{R} < 1.0085$). Among all models, only Michaelis-Menten presented one divergent transition after warmup. This single divergence from the 5000 iterated samples does not interfere with the Bayesian inference of the model.

Only the Weibull growth model and Logistic growth function presented the n_{eff}/N ratio lower than 0.10 for at least one of the parameters estimated in the model. This may indicate some autocorrelation in the MCMC sampling after the warmup period. Considering 460 estimated parameters per model, only a single parameter for two models presented a low n_{eff}/N ratio, they were: the Weibull growth model $\tau_{\theta_{2|jk}} = \tau_{\kappa_{jk}}$, and Logistic growth function $\tau_{\theta_{3|jk}} = \tau_{\beta_{jk}}$ parameter. Therefore, these models were refitted to the data with a slightly

different MCMC (HMC) sampler setup by setting the number of thinning to two, increasing the iteration size to 10,000, and keeping the warm-up period at 50% of the sample, to ensure that at the end MCMC the total post-warmup draws was the same as all compared models. Thus, the n_{eff}/N ratio was above 0.10 for these parameters, as a reference described in the diagnosis for the model.

3.3. After fitting the model

3.3.1. Comparison and models selection

Table 4 shows the comparative indices for the Hierarchical Bayesian nonlinear models fitted to the Pacific white shrimp growth data. The difference in ELPD computed through the LOO cross-validation between the second-best model (Michaelis-Menten curve) was -18.8 (9.8 standard error), which is a relevant difference.

Table 4: Comparative indices for the Bayesian Hierarchical nonlinear models fitted to growth data of the Pacific white shrimp farming (*Litopenaeus vannamei*). Values in parentheses represent the standard error of the estimates

Model	**LOO-IC[1]**	**WAIC[2]**	**Difference[4] ELPD[3]with LOO**	**Effective number of parameters**
Weibull growth	2705.2 (80.1)	2661.3 (77.4)	0.0 (0.0)	228.2 (16.2)
Michaelis-Menten	2742.8 (79.5)	2702.1 (77.1)	-18.8 (9.9)	259.0 (15.9)
MMF	3134.2 (80.9)	3109.0 (79.2)	-214.5 (21.6)	230.8 (15.0)
von Bertalanfy	3472.0 (74.5)	3432.9 (72.5)	-383.4 (21.2)	247.9 (14.1)
Logistic function	4320.3 (54.1)	4228.8 (49.6)	-807.6 (30.9)	327.5 (14.6)
Gompertz function	4777.8 (71.)	4769.0 (70.4)	-1036.3 (37.3)	187.7 (10.8)

[1] LOO-CI is Information Criterion of Leave-one-out cross-validation
[2] WAIC is Watanabe-Akaike information criterion
[3] ELPD is the theoretical expected log pointwise predictive density for a new dataset
[4] The difference in ELPD computed with LOO, relative to the highest ELPD model

3.3.2. Diagnostic and assessment of the selected model

Figures 2, 3, and 4 show the $PPC_{Posterior}$ analysis from the Weibull Bayesian Hierarchical model fitted to Pacific white shrimp growth data, at the population level, pond (tank) group, and at production cycle group (nested to the pond), respectively. It can be seen that the density plot of observed weight w (dark curve) is relatively superimposed on simulations drawn from the posterior predictive distribution w_{rep} (thin, lighter lines). This overlap is very apparent for the model at the level of the production cycle (Figure 4) and at the level of the pond group (Figure 3), but at the population level, we observe some misalignment. Such a result may indicate that the model is misspecified at the population level (Gelman et al., 2004).

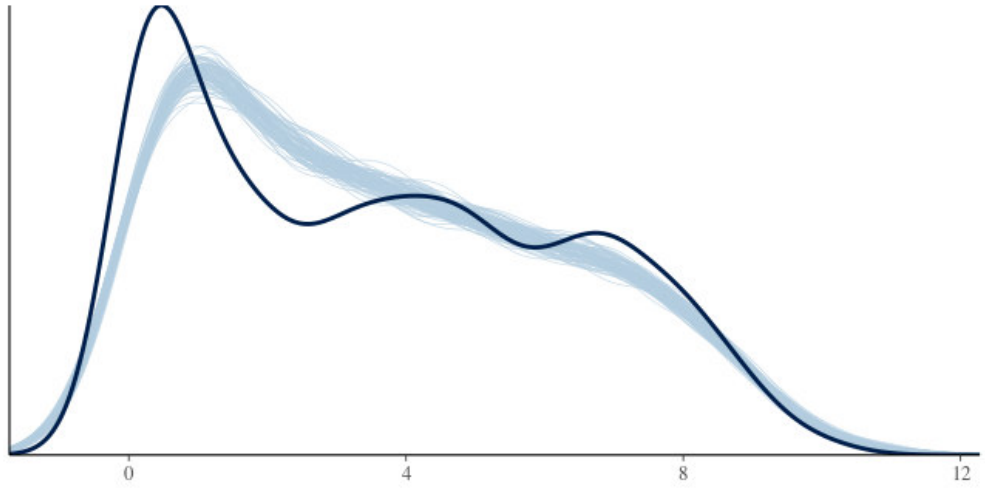


Figure 2: Posterior predictive check ($PPC_{posterior}$) for population level from Weibull Bayesian Hierarchical model fitted to Pacific white shrimp (*Litopenaeus vannamei*) growth data. Density plot of observed weight w (dark curve), and simulations drawn from the posterior predictive distribution of reproductive weight w_{rep} (thin, lighter lines)

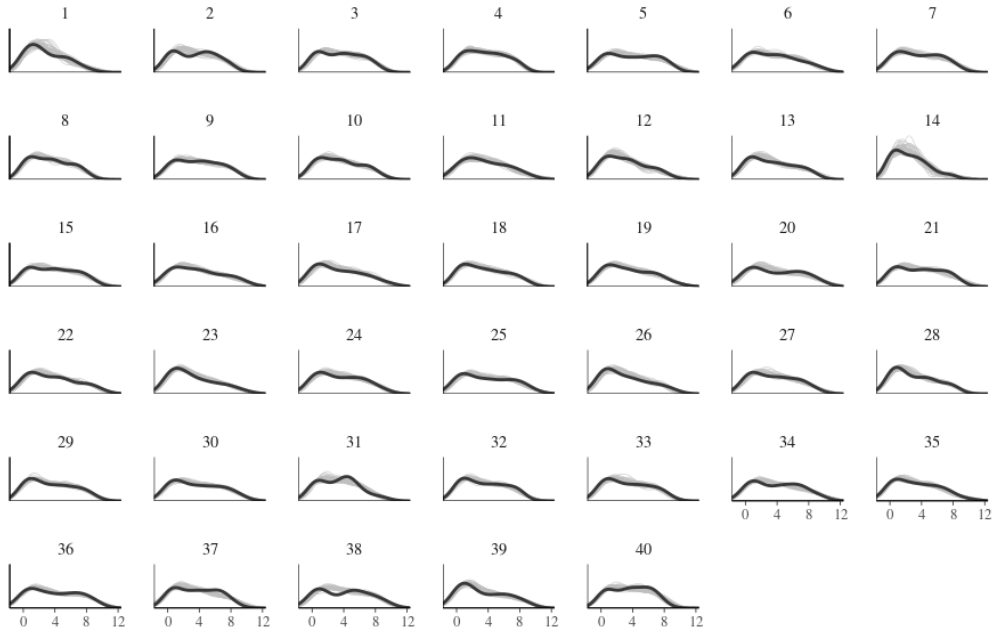


Figure 3: Posterior predictive check ($PPC_{posterior}$) for pond group (tank levels), from Weibull Bayesian Hierarchical model fitted to Pacific white shrimp (*Litopenaeus vannamei*) growth for data from a farm in northeastern Brazil. Density plot of observed weight w (dark curve), and simulations drawn from the posterior predictive distribution of reproductive weight w_{rep} (thin, lighter lines)

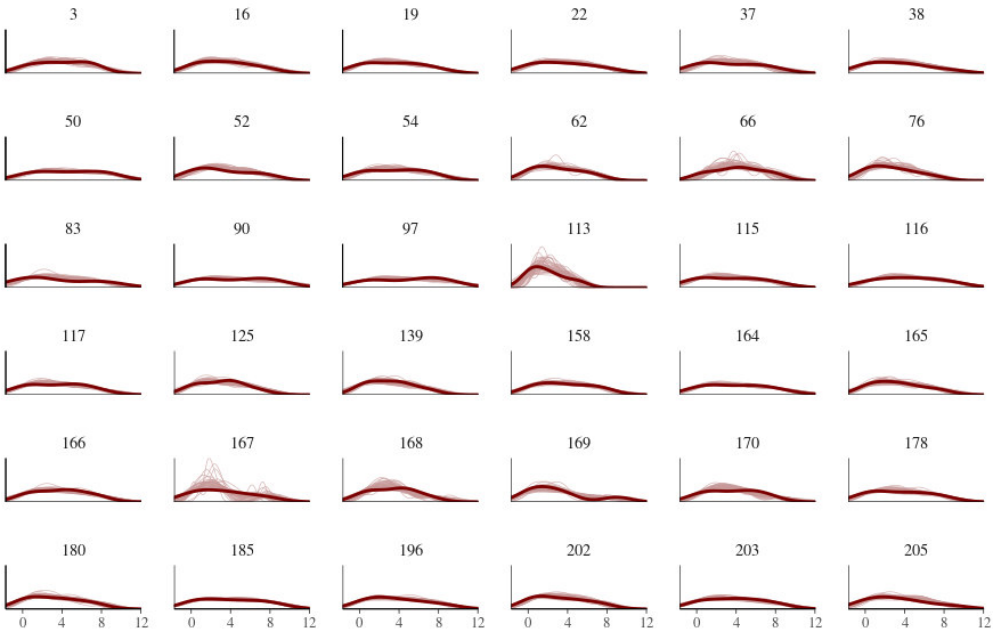


Figure 4: Posterior predictive check ($PPC_{posterior}$) within production cycles hierarchical levels of 36 production cycles nested with pond group (chosen randomly), from Weibull Bayesian Hierarchical model fitted to Pacific white shrimp (*Litopenaeus vannamei*) growth for data from a farm in northeastern Brazil. Density plot of observed weight w (dark curve), and simulations drawn from the posterior predictive distribution of reproductive weight w_{rep} (thin, lighter lines)

Parameter estimates for the Bayesian hierarchical Weibull growth model fitted to shrimp (*Litopenaeus vannamei*) data from a farm in northeastern Brazil are shown in Table 5.

Table 5: Parameter estimates of the Bayesian hierarchical Weibull growth model fitted to shrimp (*Litopenaeus vannamei*) growth data from a farm in northeastern Brazil

Hierarchical level	**Parameter**	**Mean Estimation**	**Standard deviation**	**Low-CI 95%[1]**	**Up-CI 95%[1]**
Population	α_i	79.40	5.41	68.72	90.02
	β_{ijk}	3.53	0.26	3.06	4.07
	κ_{ijk}	1.41	0.01	1.39	1.44
	σ_i	0.52	0.01	0.50	0.54
Pond	$\tau_{\beta_{jk}}$	$4.00 \cdot 10^{-4}$	$2.99 \cdot 10^{-4}$	$1.57 \cdot 10^{-5}$	$1.11 \cdot 10^{-3}$
	$\tau_{\kappa_{jk}}$	$1.92 \cdot 10^{-2}$	$5.28 \cdot 10^{-3}$	$6.99 \cdot 10^{-3}$	$2.84 \cdot 10^{-2}$
Cycle-Pond	τ_{β_k}	$3.94 \cdot 10^{-4}$	$3.01 \cdot 10^{-4}$	$1.52 \cdot 10^{-5}$	$1.11 \cdot 10^{-3}$
	τ_{κ_k}	$5.33 \cdot 10^{-2}$	$2.15 \cdot 10^{-3}$	$4.93 \cdot 10^{-2}$	$5.77 \cdot 10^{-2}$

[1] The lower boundary of 95% credibility interval
[2] The upper boundary of 95% credibility interval

3.4. Validating the selected model

Table 6 shows the correctly (True) and wrongly (False) predicted values within the credibility interval (95%) from the Posterior Predictive Distribution of the Hierarchical Weibull growth model, both at the pond group level and at the production cycle level. Figure 5 shows the results of predictive accuracy for slaughter weight, 3 biometrics (21 days) before harvesting. Figure 6 shows the accuracy results of pond group-level for production cycles completely new to the model.

Table 6: The estimated values of the prediction accuracy of the Weibull Hierarchical Growth Model for the pond group level and production cycle. TRUE means that the prediction values are within the credibility interval (95%) of the a posteriori predictive distribution, and FALSE means that the true values are outside the interval. The results were divided by the total number of predictions made to present as a percentage format

Level		**Data**	
		True	False
Model	Pond	95.76%	4.24%
	Cycle	85.71%	14.29%

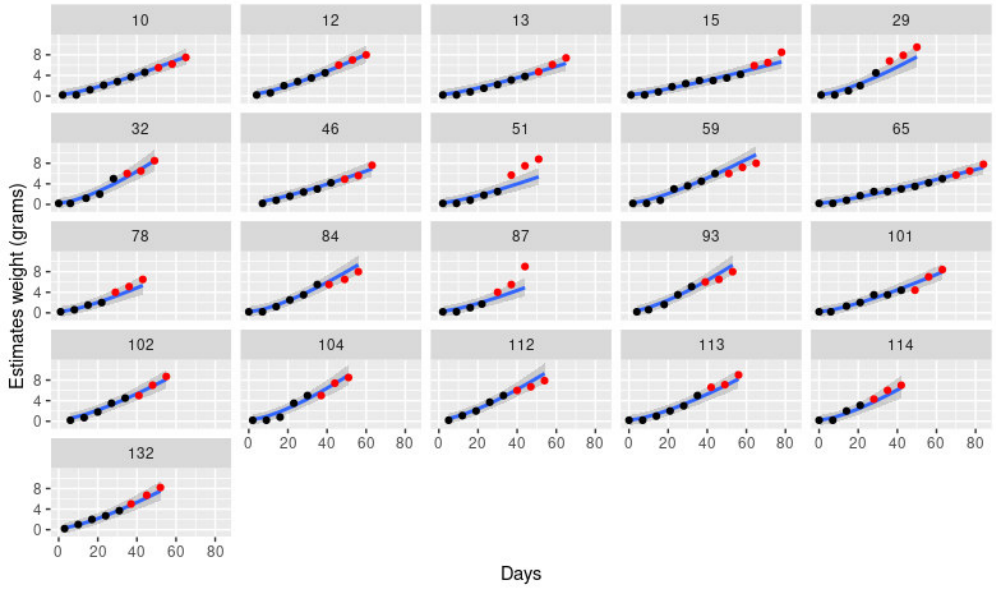


Figure 5: Prediction of weight (g) of shrimp (*Litopenaeus vannamei*) over time (days) from the Weibull Hierarchical Bayesian model at the production cycle level (cycle number: 10, 12 ,13, 15, 29, 32, 46, 51, 59, 65, 78, 84, 87, 93, 101, 102, 104, 112, 113, 114, and 132), for new data (data not used in model learning) considering 3 biometrics before harvesting (red points) for Pacific white shrimp growth farmed in northeastern Brazil

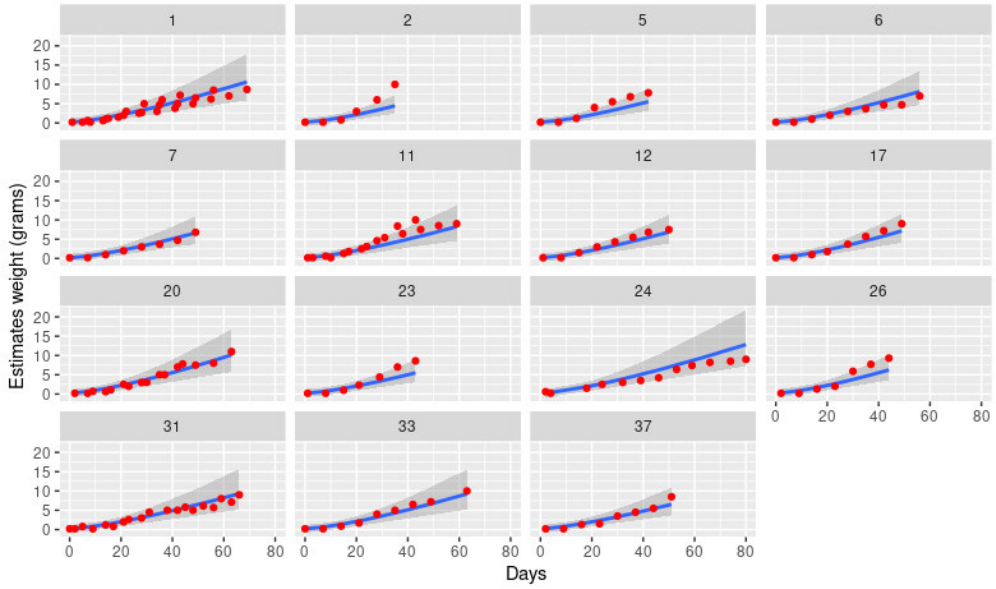


Figure 6: Prediction of weight (g) of shrimp (*Litopenaeus vannamei*) over time (days) from the Weibull Hierarchical Bayesian model at the pond group level, for new data (data not used in model learning) for Pacific white shrimp growth farmed in northeastern Brazil

3.5. Model sensitivity in detecting treatment differences

Figure 7 shows the estimates of the kappa parameter ($\theta_{2|k} = \kappa_k$) of the Weibull Hierarchical Bayesian model of growth for the comparison at the production cycle level between batches produced in the shrimp farm in northeastern

Brazil. The batches were grouped (Table 3) by the similarity through the traditional zootechnical performance indexes calculated on the farm. The proposed method detected a significant difference between the estimates of the kappa parameter compared within the group. Production cycle 47 stood out from cycle 204 in group 1. In group 2, cycles 180 and 80 were superior to cycles 61 and 172, as well as the other batches that can be noticed in the other groups in Figure 7. Due to the absence of an experimental design, the lack of information about the management of the productive environment, and the nonexistence of other variables in the provided database, it was not possible to determine the reason for this superiority between the cycles detected by the method.

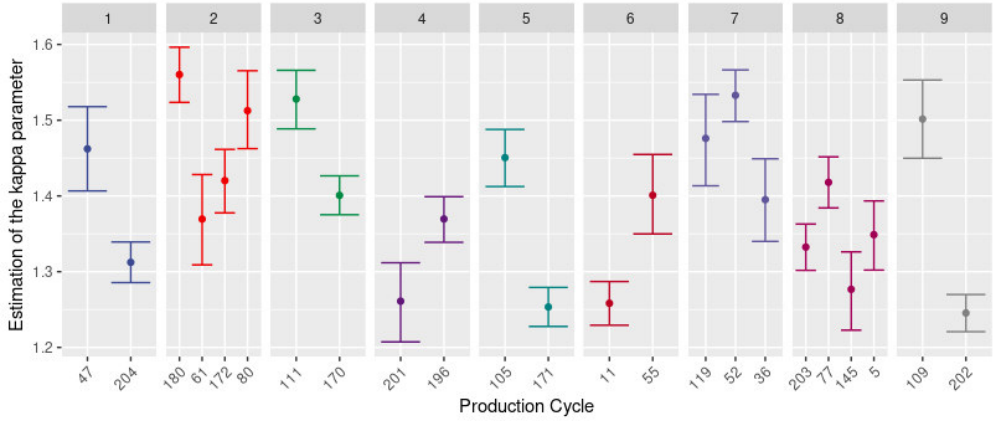


Figure 7: Production cycle of Pacific white shrimp (*Litopenaeus vannamei*) culture in a shrimp farm in northeastern Brazil, grouped (1 to 9) by the similarity of zootechnical performance for sensitivity analysis in the treatment comparison through the proposed Weibull Hierarchical Bayesian model. Comparison of productive cycles (x-axis) with kappa ($\theta_{2|k} = \kappa_k$) parameter estimate (y-axis) and 95% credibility interval estimate for each comparative cycle

4. Discussion

4.1. Before fitting a model

4.1.1. Prior predictive checking

In the quest to find a good model to represent shrimp growth in an industrial shrimp farming production system, it is of paramount importance to elicit reasonable prior probability distributions in the Bayesian context. The PPC_{Prior} feature provides a way to elicit the prior distributions and also refine the model without using the data (Gelman et al., 2020), that is, it provides a prior distribution understanding in the context of a generative model (Box, 1980; Gabry et al., 2019). Gelman et al. (2017) state that the prior marginal distribution for the data reflects the interplay between the prior distribution on the parameters and the likelihood, and assessing them through the PPC_{Prior} is a good way to understand how prior distributions actually work for a given problem.

It is important to note, for all nonlinear growth models, the $\theta_{1|i} = \alpha_i$ parameter was considered only at the population level in the Bayesian Hierarchical

structure. This parameter is related to the theoretical asymptotic weight intrinsic to the species or the individual's genetics. In the case of incomplete or limited data, usually present in aquaculture, the prior density function for this parameter was based on wild animals (Zarzar et al., 2022). Its true value in aquaculture is unknown and difficult to infer. However, it is more interesting from a producer's point of view to understand the behavior of the weight over time during the period the animal remains on the farm than the theoretical weight at the limit of the time to infinity, which is unattainable in practice. Therefore, in order to regularize the bias caused by the incomplete or limited data, the priori density distributions for this parameter were very informative with a mean of around 85 and standard deviation 1 to 5. Furthermore, this model structure in which the $\theta_{1|i} = \alpha_i$ parameter is only at the population level contributes to the comparative analysis between different production cycles or different tanks by comparing the other parameters (e.g. $\theta_{2|ijk} = \kappa_{ijk}$ and $\theta_{3|ijk} = \beta_{ijk}$) (see Section 4.5), as it also regularizes this comparison in cases where the strong correlation between parameters in nonlinear models is greater.

4.2. Fitting a model

4.2.1. Health Diagnosis and quality of the MCMC chains convergences

After elicitation of the prior probability distributions, an important step is to adjust the candidate models for modeling. Ensuring the quality and health of MCMC chains in Bayesian modeling in numerical integration methods as in this research, allows good inferences of the posterior distribution, and consequently selection of reliable models.

4.3. After fitting the model

4.3.1. Comparison and models selection

Once candidate models are fitted to the dataset, comparative indexes (Table 4) will judge which candidate models are the best. We define in this research as "the best model", not only when it is well adjusted to the observed data set, but mainly well adjusted to the studied phenomenon. Therefore, such defined selection indexes reflect how well the model can predict new data not trained by the model, that is, how flexible the model is, in light of new information. In other words, the indexes estimate the predictive precision of the model (accuracy), or the expected prediction error, depending on the point of view.

Among all candidates, the Weibull growth model was selected based on the ELPD difference, since the other indices in Table 4 were relatively close to the second-best model (Michaelis-Menten). According to Sivula et al. (2020), an ELPD-LOO difference greater than 4 may be relevant in comparing models and also should take into account the standard error estimate of the difference ELPD-LOO between proposed models in Bayesian models comparison and selection.

4.3.2. Diagnostic and assessment of the selected model

In short, The $PPC_{Posterior}$ is the feature assessing if the fitted model provides valid predictions about reality, in which, through simulated replicated data under the generative model are compared to the actual observed data (Gelman and Hill, 2006). The most interesting analysis of the posterior predictive distribution is that it represents the uncertainty over two different causes: the one associated with the estimated parameters of the posterior distribution, and the other one regarding the sampling distribution or the assumed probabilistic model for the data.

Although posterior predictive checking makes use of the data twice, first for the fitting and second for the checking (Gabry et al., 2019), it allows for detecting conflicts between the data and the model. Therefore, it is not interesting to make inferential decisions based on this analysis. For this reason, its usage was limited to discrepancy measures to study model adequacy, not for model comparison and inference (Meng, 1994). For comparative purposes between different models, the leave-one-out cross-validation (LOO-CV) approximated by Pareto smoothed importance sampling (PSIS) method, and WAIC was carried out to avoid the double use of data.

Even with indications that the model is misspecified at the population level, excellent diagnoses were observed at the pond level and production cycle level. Considering that the main interest of a farm is to evaluate the performance of its pond (tank) and mainly make predictions about its production cycles, we concluded that the model can be useful for such hierarchical levels as they are well specified. The $PPC_{Posterior}$ analysis evidence that the Weibull Bayesian hierarchical growth model is able to simulate new data similar to the observed values over the groups' levels for the phenomenon studied.

4.4. Validating the selected model

The estimated forecast accuracy values of the Weibull Hierarchical Growth model for pond level and production cycle level were reasonable for application in shrimp farm management systems. Obviously, the accuracy of this model increases as more information makes up the training model and, consequently, the closer the future prediction that we want to perform until harvesting. Therefore, the accuracy in forecasting shrimp slaughter weight can be improved by considering two or one biometrics before harvesting (Figure 5).

It is worth mentioning that in the predictions at the production cycle level, the model considers the time series information on previous biometrics throughout the cultivation (black dots, Figure 5) to predict the future (red dots, Figure 5). Predicting a new cultivation without any previous information is more challenging for any model (Figure 6). However, for the Bayesian Hierarchical models, this is possible due to the sharing of statistical information (borrowing strength) among the levels of the hierarchy and the dependence of information among the levels of the specific group (Xu et al., 2020). This means that previous production cycles carried out in the specific tank have a relative dependence that helps to compose the information to predict a future new production cycle not yet started. In addition, growth curves from other ponds also contain

information that contributes to the same purpose. In this way, the proposed method has inference benefits when compared to growth modeling proposals usually practiced in aquaculture that assume independence among production cycles and among tanks.

4.5. Model sensitivity in detecting treatment differences

The kappa parameter for the Weibull nonlinear growth model is directly related to the animal's growth rate. The sensitivity analysis, compares it with traditional zootechnical indexes commonly used in shrimp farms, observing whether they are able to identify subtle differences between production cycles.

The results showed that even with very similar zootechnical performance between one lot and another (Tabela 3), they are different production cycles, cultivated in different periods, often with different initial stocking densities, and produced in tanks with different sizes. These differences reflect on the growth rate of the animals but are not detected with traditional indices. However, the proposed method and model made it possible to detect these subtle differences (Figure 7) between the production cycles classified as similar by the farm manager.

The proposed method becomes more sensitive because the traditional analysis does not take into account the information on the growth curve as a whole (continuous information), but only discrete punctual information (weekly) until harvesting, reducing the sensitivity of the analysis. Due to the absence of an experimental design, the lack of information about the management of the productive environment, and the nonexistence of other variables in the provided database, it was not possible to determine the reason for this superiority between the cycles detected by the method.

Therefore, the sensitivity of the proposed method can collaborate as an additional tool in the investigative advancement of the scientific and business community in technological development in the area of management, nutrition, and genetics, among other applications together with the design experiment framework.

5. Conclusion

Among several nonlinear models fitted to a dataset of a real production environment in northeastern Brazil, the Weibull growth equation stands out as the best, considering the Bayesian Hierarchical structure for modeling the growth of the Pacific white shrimp (*Litopenaeus vannamei*). Although the model fit diagnosis indicates that the model is misspecified at the population level, at the pond (tank), and production cycle level we obtained excellent results, concluding that the model was considered well-adjusted for such hierarchical levels. Thus, the finding of this research is useful as these groups' levels are the main interest in real-world analysis and management within the shrimp farm.

Results showed greater sensitivity in detecting differences between possible comparative treatments. Therefore, this tool can be used to improve processes

and products within a production environment focused on modern Industry 4.0 in aquaculture.

Studies are suggested in order to improve the fit of the model at the population level, and the modeling of animal size variability over time to improve the model. Further studies are recommended in controlled laboratory environments for an accurate conclusion on the sensitivity comparative method with an interest in shrimp growth comparison. In addition, it is suggested to investigate this tool applied to any other aquaculture organisms and even outside it, which share the same characteristics of incomplete or limited data.

References

Abinaya, N., Susan, D., Kumar, R., 2021. Naive bayesian fusion based deep learning networks for multisegmented classification of fishes in aquaculture industries. Ecological Informatics 61, 101248.

Aragón-Noriega, E.A., Mendivil-Mendoza, J.E., Alcántara-Razo, E., Valenzuela-Quiñónez, W., Félix-Ortiz, J.A., 2017. Multi-criteria approach to estimate the growth curve in the marine shrimp, penaeus vannamei boone, 1931 (decapoda, penaeidae). Crustaceana 90, 1517–1531.

Bailey, J., Alanärä, A., 2001. A test of a feed budget model for rainbow trout, oncorhynchus mykiss (walbaum). Aquaculture Research 32, 465–469.

Bates, D., Watts, D., 2007. Nonlinear regression analysis and its applications, 2nd.

Box, G.E., 1980. Sampling and bayes' inference in scientific modelling and robustness. Journal of the Royal Statistical Society: Series A (General) 143, 383–404.

Cacho, O.J., Kinnucan, H., Hatch, U., 1991. Optimal control of fish growth. American Journal of Agricultural Economics 73, 174–183.

Chiu, W.A., Wright, F.A., Rusyn, I., 2017. A tiered, bayesian approach to estimating of population variability for regulatory decision-making. Altex 34, 377.

Cho, C.Y., Bureau, D.P., 1998. Development of bioenergetic models and the fish-prfeq software to estimate production, feeding ration and waste output in aquaculture. Aquatic Living Resources 11, 199–210.

Dumas, A., Dijkstra, J., France, J., 2008. Mathematical modelling in animal nutrition: a centenary review. The Journal of Agricultural Science 146, 123–142.

Einen, O., Holmefjord, I., Åsgård, T., Talbot, C., 1995. Auditing nutrient discharges from fish farms: theoretical and practical considerations. Aquaculture Research 26, 701–713.

Estrada-Pérez, A., Ruiz-Velazco, J.M., Hernández-Llamas, A., Zavala-Leal, I., Martínez-Cárdenas, L., 2016. Deterministic and stochastic models for analysis of partial harvesting strategies and improvement of intensive commercial production of whiteleg shrimp (litopenaeus vannamei). Aquacultural engineering 70, 56–62.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., Gelman, A., 2019. Visualization in bayesian workflow. Journal of the Royal Statistical Society: Series A (Statistics in Society) 182, 389–402.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian data analysis. Chapman and Hall/CRC.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian Data Analysis. 2nd ed. ed., Chapman and Hall/CRC.

Gelman, A., Hill, J., 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge university press.

Gelman, A., Hwang, J., Vehtari, A., 2014. Understanding predictive information criteria for bayesian models. Statistics and computing 24, 997–1016.

Gelman, A., Rubin, D.B., 1992. lnference from iterative simulation using multiple sequences. Statistical Science 7, 457–472. doi:10.1214/ss/1177011136.

Gelman, A., Simpson, D., Betancourt, M., 2017. The prior can often only be understood in the context of the likelihood. Entropy 19, 555.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C.C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.C., Modrák, M., 2020. Bayesian Workflow. arXiv e-prints , arXiv:2011.01808arXiv:2011.01808.

Hoffman, M.D., Gelman, A., 2014. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. Journal of Machine Learning Research 15, 1593–1623. URL: jmlr.csail.mit.edu/papers/v15/hoffman14a.html.

Kruschke, J., 2014. Doing bayesian data analysis: A tutorial with r, jags, and stan .

Martinez, J.A., Seijo, J.C., 2001. Economics of risk and uncertainty of alternative water exchange and aeration rates in semi-intensive shrimp culture systems. Aquaculture Economics & Management 5, 129–145.

McElreath, R., 2020. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC.

de Melo Filho, M.E.S., Owatari, M.S., Mouriño, J.L.P., Carciofi, B.A.M., Soares, H.M., 2020. Empirical modeling of feed conversion in pacific white shrimp (litopenaeus vannamei) growth. Ecological Modelling 437, 109291.

Meng, X.L., 1994. Posterior predictive p-values. The annals of statistics 22, 1142–1160.

Miguez, F., Archontoulis, S., Dokoohaki, H., 2018. Nonlinear regression models and applications. Applied statistics in agricultural, biological, and environmental sciences , 401–447.

Neal, R.M., 2011. MCMC using Hamiltonian dynamics, in: Handbook of Markov chain Monte Carlo.. Boca Raton, FL: CRC Press, pp. 113–162.

Pache, M.C.B., Sant'Ana, D.A., Rezende, F.P.C., de Andrade Porto, J.V., Rozales, J.V.A., de Moraes Weber, V.A., Junior, A.d.S.O., Garcia, V., Naka, M.H., Pistori, H., 2022. Non-intrusively estimating the live body biomass of pintado real® fingerlings: A feature selection approach. Ecological Informatics 68, 101509.

Piironen, J., Vehtari, A., et al., 2017. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electronic Journal of Statistics 11, 5018–5051.

Pinheiro, J., Bates, D., 2006. Mixed-effects models in S and S-PLUS. Springer science & business media.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Ruiz-Velazco, J.M., González-Romero, M.A., Estrada-Perez, N., Hernandez-Llamas, A., 2021. Evaluating partial harvesting strategies for whiteleg shrimp litopenaeus (penaeus) vannamei semi-intensive commercial production: profitability, uncertainty, and economic risk. Aquaculture International , 1–13.

Ruiz-Velazco, J.M., Hernández-Llamas, A., Gomez-Muñoz, V.M., 2010. Management of stocking density, pond size, starting time of aeration, and duration of cultivation for intensive commercial production of shrimp litopenaeus vannamei. Aquacultural Engineering 43, 114–119.

Sanchez, I., Gonzalez, I., 2021. Monitoring shrimp growth with control charts in aquaculture. Aquacultural Engineering 95, 102180.

Sivula, T., Magnusson, M., Vehtari, A., 2020. Uncertainty in bayesian leave-one-out cross-validation based model comparison. arXiv preprint arXiv:2008.10296 .

Stan Development Team, 2020. RStan: the R interface to Stan. URL: `http://mc-stan.org/`. r package version 2.21.2.

Tian, X., Leung, P., Hochman, E., 1993. Shrimp growth functions and their economic implications. Aquacultural Engineering 12, 81–96.

Vehtari, A., Gelman, A., Gabry, J., 2017. Practical bayesian model evaluation using leave-one-out cross-validation and waic. Statistics and computing 27, 1413–1432.

Watanabe, S., Opper, M., 2010. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. Journal of machine learning research 11, 3571–3594. URL: www.jmlr.org/papers/v11/watanabe10a.html.

Xu, G., Zhu, H., Lee, J.J., 2020. Borrowing strength and borrowing index for bayesian hierarchical models. Computational statistics & data analysis 144, 106901.

Yu, R., Leung, P., 2010. A bayesian hierarchical model for modeling white shrimp (litopenaeus vannamei) growth in a commercial shrimp farm. Aquaculture 306, 205–210.

Yu, R., Leung, P., Bienfang, P., 2006. Predicting shrimp growth: artificial neural network versus nonlinear regression models. Aquacultural Engineering 34, 26–32.

Zarzar, C.A., Silva, E.M., Fernandes, T.J., De Oliveira, I.R.C., 2022. Evidence of parameters underestimation from nonlinear growth models for data classified as limited. Computers and Electronics in Agriculture 200, 107196.

Zhang, Y.D., Naughton, B.P., Bondell, H.D., Reich, B.J., 2020. Bayesian regression using a prior on the model fit: The r2-d2 shrinkage prior. Journal of the American Statistical Association , 1–13.