

UNIVERSIDADE FEDERAL DE VIÇOSA

**Modelagem espacial do desmatamento no sul do Amazonas: uma abordagem
utilizando análise de sobrevivência**

Mauricio dos Anjos da Silva
Magister Scientiae

**VIÇOSA - MINAS GERAIS
2025**

MAURICIO DOS ANJOS DA SILVA

Modelagem espacial do desmatamento no sul do Amazonas: uma abordagem utilizando análise de sobrevivência

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

Orientador: Nerilson Terra Santos

Coorientador: Sebastiao Martins Filho

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

S586m
2025
Silva, Maurício dos Anjos da, 1999-
Modelagem espacial do desmatamento no sul do
Amazonas: uma abordagem utilizando análise de sobrevivência /
Maurício dos Anjos da Silva. – Viçosa, MG, 2025.
1 dissertação eletrônica (115 f.): il. (algumas color.).

Inclui apêndices.

Orientador: Nerilson Terra Santos.

Dissertação (mestrado) - Universidade Federal de Viçosa,
Departamento de Estatística, 2025.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2025.293>

Modo de acesso: World Wide Web.

1. Análise de sobrevivência (Biometria). 2. Teoria bayesiana de decisão estatística. 3. Desmatamento - Amazônia - Métodos estatísticos. I. Santos, Nerilson Terra, 1966-. II. Universidade Federal de Viçosa. Departamento de Estatística. Programa de Pós-Graduação em Estatística Aplicada e Biometria. III. Título.

CDD 22. ed. 519.54

MAURICIO DOS ANJOS DA SILVA

Modelagem espacial do desmatamento no sul do Amazonas: uma abordagem utilizando análise de sobrevivência

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 19 de fevereiro de 2025.

Assentimento:

Mauricio dos Anjos da Silva
Autor

Nerilson Terra Santos
Orientador

Essa dissertação foi assinada digitalmente pelo autor em 15/05/2025 às 16:16:13 e pelo orientador em 16/05/2025 às 08:42:02. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse <https://siadoc.ufv.br/validar-documento>. No campo 'Código de registro', informe o código **T2NA.ZXWR.MIIS** e clique no botão 'Validar documento'.

AGRADECIMENTOS

Primeiramente a Deus.

Aos meus pais, Levy de Araújo e Sandra Aparecida, por sempre me apoiarem e por todo incentivo que depositaram em mim durante esse processo. O grande exemplo de caráter, força e resiliência que representam para mim me inspira todos os dias e me faz continuar mesmo diante de adversidades.

A meus irmãos Maurílio, Marcos e Michel por todos os momentos de alegria que me fizeram sentir melhor durante todo esse período. Com vocês compartilho minhas melhores lembranças.

Aos meus orientadores Prof. Dr. Nerilson Terra Santos e Prof. Dr. Sebastião Martins Filho, por todo acompanhamento e parceria. Sou grato por tê-los como grandes referências em minha formação, e por cada conselho que me fizeram enxergar um mundo melhor.

Ao discente Marciel Lelis Duarte, por toda paciência e conhecimento compartilhado. Sou grato por sua valiosa contribuição para a minha formação, e por poder tê-lo como uma grande referência.

A todos os meus amigos e colegas do programa. Sou grato pelos grandes momentos de alegria e carinho. Tenho cada um como um exemplo de caráter e sinto profunda admiração. A forma como me acolheram me surpreendeu, por isso sou grato a cada um de vocês.

Em especial, a meus amigos Bia, Lucas Belo, Lucas Coelho, Renata e Verônica, por todos os momentos de dificuldades que, juntos, conseguimos vencer. Sou grato a vocês por cada momento e conversa descontraída, pela grande parceria, e por me mostrarem que a melhor maneira de vencer os problemas é com um sorriso no rosto.

Ao professor Fernando Souza Bastos, pelo aceite em compor a minha banca de monografia, e pelas grandes contribuições para o trabalho. Agradeço também por todos os valiosos conselhos que me fizeram chegar até aqui. Ao secretário do Programa de Pós Graduação em Estatística Aplicada e Biometria (PPESTBIO), Júnior José Pires, por toda atenção e paciência durante todo o período de execução desse trabalho. A sua disponibilidade de sempre ajudar configura um ambiente acolhedor com relação a todos os assuntos pertinentes e que dizem respeito às dependências do programa.

A todos os professores do Departamento de Estatística pela excelente atuação durante as disciplinas que me proporcionaram grandes aprendizados para a minha formação. Agradeço pela disposição de sempre em ajudar a superar qualquer que seja a dificuldade. A todos vocês expresso minha gratidão.

À Universidade Federal de Viçosa, pela grande oportunidade de realizar a pós-graduação.

Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

A todos os amigos que contribuíram de alguma forma durante todo o processo de realização desse trabalho, os meus sinceros agradecimentos.

“Aquilo que não me destrói, me fortalece.”
(Friedrich Nietzsche)

RESUMO

SILVA, Mauricio dos Anjos da, M.Sc., Universidade Federal de Viçosa, fevereiro de 2025. **Modelagem espacial do desmatamento no sul do Amazonas: uma abordagem utilizando análise de sobrevivência.** Orientador: Nerilson Terra Santos. Coorientador: Sebastiao Martins Filho.

O desmatamento é um problema de grande interesse ambiental, e, uma vez conhecidos os seus principais índices e tendências, é possível determinar e redefinir as medidas de preservação direcionadas à degradação florestal. Por esse motivo, entender como algumas covariáveis se relacionam com o desmatamento é de grande interesse prático. Os modelos estatísticos que utilizam a abordagem bayesiana acompanham o avanço teórico e tecnológico, uma vez que dependem fortemente de recursos computacionais para lidar com diferentes distribuições a posteriori. Inserir também os efeitos aleatórios com base na dependência espacial dos dados é especialmente relevante, à medida que os dados são espacialmente correlacionados. No capítulo 1, é apresentada uma revisão teórica das principais técnicas de análise de sobrevivência: Métodos não paramétricos, paramétricos e Modelos de fragilidade, além dos modelos autorregressivos ICAR. No capítulo 2, são apresentados os passos para levantamento, coleta e análise de dados de desmatamento no sul do Amazonas utilizando a análise bayesiana e os modelos de fragilidade. Os modelos de fragilidade compartilhada são uma importante classe de modelos que permitem a inserção de termos de efeito aleatório associados a grupos de interesse. Os dados de desmatamento foram coletados anualmente a partir do ano de 1987 a 2023, por meio do site do MapBiomas e agregados em conjunto às covariáveis distância do pixel a rodovias, distância à hidrovias, distância a pontos de mineração, pastagem, população, potencialidade agrícola, terras indígenas e unidades de conservação, disponíveis no site do IBGE e MapBiomas. A partir da aplicação das técnicas de análise de sobrevivência, aliadas a uma avaliação espacial do desmatamento, foi possível detalhar o efeito de cada covariável inserida no modelo. Com isso, os modelos de análise de sobrevivência com fragilidade representam uma técnica robusta na análise do desmatamento.

Palavras-chave: Modelos de fragilidade compartilhada ; Modelagem Bayesiana; Modelo ICAR; Análise de vizinhança; Floresta Amazônica

ABSTRACT

SILVA, Mauricio dos Anjos da, M.Sc., Universidade Federal de Viçosa, February, 2025. **Spatial modeling of deforestation in southern Amazonas: an approach using survival analysis.** Adviser: Nerilson Terra Santos. Co-adviser: Sebastiao Martins Filho.

Deforestation is an issue of great environmental interest, and once its main indices and trends are known, it becomes possible to determine and redefine preservation measures aimed at mitigating forest degradation. For this reason, understanding how certain variables relate to deforestation is of significant practical interest. Statistical models that use the Bayesian approach follow theoretical and technological advances, as they rely heavily on computational resources to handle different posterior distributions. Incorporating random effects based on the spatial dependency of the data is particularly relevant since the data are spatially correlated. In Chapter 1, a theoretical review of the main survival analysis techniques is presented: non-parametric, parametric, and frailty models, as well as ICAR autoregressive models. In Chapter 2, the steps for the collection, gathering, and analysis of deforestation data in southern Amazonas using Bayesian analysis and frailty models are outlined. Shared frailty models are an important class of models that allow for the inclusion of random effect terms associated with groups of interest. Deforestation data were collected annually from 1987 to 2023 through the MapBiomas website and were combined with covariates such as the distance of pixels to roads, waterways, and mining sites, as well as pasture, population, agricultural potential, indigenous lands, and conservation units, available from the IBGE and MapBiomas websites. By applying survival analysis techniques along with a spatial evaluation of deforestation, it was possible to detail the effect of each covariate included in the model. Thus, survival analysis models with frailty represent a robust technique for analyzing deforestation.

Keywords: Shared frailty models; Bayesian modeling; ICAR model; Neighborhood analysis; Amazon Rainforest

LISTA DE ILUSTRAÇÕES

- Figura 1 - Mapa da região de estudo delimitado na região norte do Brasil, por tipo de vegetação, com as coordenadas geográficas correspondentes. As classes de vegetação incluem Água, Campinarana, Contato, Floresta Ombrófila Aberta, Floresta Ombrófila Densa, Formação Pioneira e Savana. 43
- Figura 2 - Curvas para as probabilidades de desmatamento estimadas pelo método não paramétrico de Kaplan-Meier para: (A) Unidades de conservação, (B) Áreas urbanas, (C) Terras indígenas, (D) Potencialidade agrícola e (E) Pastagem, p-valor referente ao teste Log-rank para comparação das curvas de sobrevivência estimadas. 51
- Figura 3 - Curvas para as probabilidades de desmatamento estimadas pelo método não paramétrico de Kaplan-Meier para: (A) Distância até rodovias, (B) Distância até hidrovias, (C) Distância à mineração e (D) População, p-valor referente ao teste Log-rank para comparação das curvas de sobrevivência. 52
- Figura 4 - Resíduos de Cox-Snell para avaliação da adequação geral da distribuição a posteriori aos dados, representada pelos modelos: A) Modelo Exponencial, B) Modelo Weibull e C) Modelo Lognormal. 53
- Figura 5 - Distribuições marginais para os efeitos associados às covariáveis no modelo de fragilidade para avaliação da dispersão, posição e associação com relação aos demais coeficientes, calculadas a partir da verossimilhança associada ao modelo Weibull, PA = Potencialidade agrícola, UC = Unidades de conservação. 58
- Figura 6 - Distribuições marginais para os hiperparâmetros do modelo de fragilidade associado à distribuição Weibull para a verossimilhança e o modelo autorregressivo ICAR para modelagem do termo de efeito aleatório. 59
- Figura 7 - Efeitos aleatórios para os grupos identificados associados ao desmatamento no modelo de fragilidade para o modelo Weibull. 60
- Figura 8 - Gráfico de cotovelo (Ewbow Graph) para diferentes valores de k, com representação de 2 a 15 grupos. 71
- Figura 9 - Silhueta média para diferentes valores de k, com variação para o número de grupos de 2 a 15 grupos. 72
- Figura 10 - Gráfico de silhueta para k = 5 grupos, com representação individual da silhueta média dentro de cada grupo. 73
- Figura 11 - Resíduos de Cox-Snell para verificação do ajuste do modelo paramétrico Weibull com representação de: (a) comparação do ajuste paramétrico com o estimador não paramétrico de Kaplan-Meier, e (b) comparação dos resíduos obtidos com a distribuição exponencial padrão. 81

LISTA DE TABELAS

Tabela 1 - Principais modelos paramétricos de sobrevivência com identificação de sua função de densidade de probabilidade e função de sobrevivência associada.	23
Tabela 2 – Tabela para identificação e interpretação do fator de Bayes.	31
Tabela 3 - Regras de reclassificação do desmatamento anual, com identificação da classe referente à áreas desmatadas e não desmatadas.	44
Tabela 4 - Tabela para a comparação dos modelos com o termo de fragilidade e sem o termo de fragilidade.	54
Tabela 5 - Coeficientes do modelo de fragilidade ajustado utilizando a distribuição Weibull para a verossimilhança, juntamente com o intervalo de credibilidade com os limites inferiores e superiores a 5% de probabilidade, o valor médio e o índice KLD (Kullback-Leibler divergence).	55
Tabela 6 - Tabela para identificação dos valores associados aos grupos identificados por meio do método K-means para o modelo de fragilidade	60
Tabela 7 - Ajustes individuais para verificação do primeiro passo do método de Collet com identificação do DIC e Fator de Bayes como critério de seleção do melhor modelo com relação à cada covariável.	74
Tabela 8 - Tabela para verificação do passo 2 no método de Collet com a comparação gerada pela retirada de cada variável com o ajuste global.	75
Tabela 9 – Tabela para a verificação do passo 3 da aplicação do método de Collet com representação do DIC e Fator de Bayes para comparar o ajuste global do modelo.	75
Tabela 10 - Tabela para verificação do passo 4 do método de Collet para validação da remoção das covariáveis no passo 2, com identificação do DIC e Fator de Bayes.	76
Tabela 11 - Tabela para verificação do passo 5 para o método de Collet para definição das covariáveis selecionadas pelo procedimento, com comparação dos modelos por meio do DIC e Fator de Bayes.	77
Tabela 12 - Coeficientes estimados para os modelos paramétricos de sobrevivência, ajustados para a modelagem do desmatamento.	79
Tabela 13 - Tabela para comparação de modelos com apresentação das métricas AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion) para os modelos paramétricos avaliados nos dados de desmatamento.	80
Tabela 14 - Coeficientes estimados a partir dos dados de desmatamentos para o modelo paramétrico Weibull.	82
Tabela 15 - Coeficientes do modelo sem o termo de fragilidade, ajustado utilizando a distribuição Weibull para a verossimilhança, com o intervalo de credibilidade com os limites inferiores e superiores a 5% de probabilidade, o valor médio.	83

LISTA DE SIGLAS E ABREVIATURAS

BF – Bayes Factor

CONAMA – Conselho Nacional do Meio Ambiente

DETER – Sistema de Detecção do Desmatamento em Tempo Real

DETER-b – Sistema de Detecção de Desmatamento em Tempo Real – Refinado para Áreas Pequenas

DIC – Deviance Information Criterion

GFW – Global Forest Watch

IBAMA – Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis

INPE – Instituto Nacional de Pesquisas Espaciais

SUMÁRIO

INTRODUÇÃO.....	13
Referências.....	16
CAPÍTULO 1: REFERENCIAL TEÓRICO.....	17
1.1. O desmatamento e a biodiversidade	17
1.2. A análise de sobrevivência.....	20
1.2.1. O estimador não paramétrico de Kaplan-Meier	21
1.2.2. Técnicas paramétricas para análise de sobrevivência.....	22
1.3. O modelo ICAR	24
1.4. Matriz de vizinhanças.....	24
1.5. O modelo de fragilidade compartilhada	25
1.5.1. Uma abordagem bayesiana	27
1.5.2. Deviance Information Criterion – DIC	29
1.5.3. O fator de Bayes	30
1.5.4. O método K-Means	31
1.5.5. Seleção de variáveis	32
Referências.....	34
CAPÍTULO 2: MODELAGEM ESPACIAL DO DESMATAMENTO NO SUL DO AMAZONAS: UMA ABORDAGEM UTILIZANDO ANÁLISE DE SOBREVIVÊNCIA.....	38
2.1.1. Resumo.....	38
2.1.2. Abstract.....	39
2.2. Introdução	40
2.3. Material e métodos.....	43

2.3.1. Área de estudo	43
2.3.2. Dados de desmatamento	44
2.3.3. Covariáveis	45
2.3.4. Métodos não paramétricos	45
2.3.5. Métodos paramétricos.....	46
2.3.6. Os modelos de fragilidade.....	47
2.3.7. Adequação do modelo de fragilidade	59
2.4. Resultados	50
2.4.1. Estimador de Kaplan-Meier	50
2.4.2. Análise da distribuição <i>a posteriori</i> a partir dos resíduos de Cox-Snell.....	52
2.4.3. Resultados para o modelo de fragilidade	54
2.4.4. Efeitos aleatórios.....	59
2.5. Discussões.....	61
2.6. Conclusões	66
Referências.....	66
APÊNDICE A - O método K-Means para agrupamento.....	70
APÊNDICE B – Seleção de variáveis.....	74
APÊNDICE C – Ajuste do modelo paramétrico	78
APÊNDICE D – Script para aplicação do método K-means.....	84
APÊNDICE E - Script para a aplicação do método não paramétrico.....	88
APÊNDICE F – Script para o ajuste do modelo paramétrico.....	95
APÊNDICE G – Script para o ajuste do modelo de fragilidade.....	103

INTRODUÇÃO

A Floresta Amazônica desempenha um papel crucial e de grande relevância no cenário mundial. Dentre esses, é possível citar o estoque de carbono, a melhoria no fluxo hidrológico e os rios voadores (Fearnside e Laurance, 2012). Além do grande efeito na temperatura global, a Floresta Amazônica possui efeito ainda mais direto em regiões vizinhas, caracterizando o microclima local e contribuindo para um fluxo de humidade e controle da temperatura dessas regiões.

Além disso, o desmatamento é um problema mundial que possui efeito direto no cenário econômico, no equilíbrio ecológico e no meio social (Monteiro et al., 2008). A retirada da floresta nativa de um local pode causar danos irreversíveis a um ecossistema, e, dentre esses, é possível citar a degradação do solo, a contribuição para mudanças climáticas e a desertificação (Fearnside, 2005).

Diversos métodos estatísticos têm sido empregados para a modelagem de padrões no desmatamento, incluindo modelos de regressão ponderados, modelos lineares generalizados como regressão logística e Poisson (Saha et al., 2020), métodos de machine learning (Dias et al., 2024; Dominguez et al., 2022), entre outros. Hargrave e Kis-Katos (2013), apresentam uma abordagem relacionando o desmatamento a aspectos econômicos. Outro estudo foi proposto por Jusys (2016), que adotou tanto fatores econômicos quanto ambientais para utilização da metodologia GWR (*Geographically Weighted Regression*). Saha et al. (2020) estudaram o desmatamento a partir de diversos fatores sociais, características físicas do local, e fatores econômicos utilizando regressão logística e florestas aleatórias (*random forest*). Esses estudos evidenciam a grande relevância de considerar fatores econômicos e sociais no estudo do desmatamento, e destacam ainda a necessidade de abordagens robustas para se estudar as suas dinâmicas (Kyere-Boateng e Marek, 2021).

O desmatamento pode ser modelado de forma a considerar o tempo até que determinada porção de área no espaço seja desmatada. Nesse sentido, ao longo do tempo estudado, uma área pode ter sido desmatada já no início do estudo, no meio desse período ou mesmo não ser desmatada.

Outro aspecto importante é que, em várias situações, podem existir covariáveis não mensuradas e importantes para o modelo, pois alteram diretamente o risco de o evento acontecer ou mesmo o tempo de ocorrência do evento

propriamente dito. Essas covariáveis representam fatores que afetam a variável resposta do estudo, mas que não é observada nos dados. Surge, portanto, a necessidade de inserir o efeito dessas covariáveis no modelo. Essa situação configura um cenário de aplicação da análise de sobrevivência, em que o tempo é uma variável de interesse, e dos modelos de fragilidade que levam em consideração os efeitos aleatórios (Balan e Putter, 2020).

A análise de sobrevivência é um importante ramo da estatística que surge a partir do interesse em modelar o efeito de covariáveis no tempo até a ocorrência de algum evento ou fator de estudo (Colosimo e Giolo, 2024). Essa abordagem é caracterizada pela presença de censuras, que são representadas por indivíduos que não experimentaram o evento de interesse por algum motivo, em que esses indivíduos são identificados por cada observação dos dados. Dessa forma, a modelagem estatística do desmatamento por meio da análise de sobrevivência pode ser conduzida considerando áreas desmatadas como evento de interesse, e áreas não desmatadas como eventos censurados. Métodos que utilizam apenas modelos que não incorporam as censuras, perdem a informação contida nesses indivíduos. Colosimo e Giolo (2024) ressaltam que, em estudos de análise de sobrevivência, todas as observações possuem informações relevantes, mesmo que parcialmente, no caso de elementos censurados. Além disso, não considerar as censuras no estudo pode levar a conclusões viesadas.

Os modelos de fragilidade representam uma importante classe de modelos na análise de sobrevivência. Sua aplicação permite a inserção de efeitos aleatórios no modelo, além dos efeitos das covariáveis (Balan e Putter, 2020; Hougaard, 1995; Colosimo e Giolo, 2024). A relevância associada a essa abordagem se encontra no fato de que, além de modelar o efeito de causas diretas do desmatamento como a expansão da fronteira agrícola, mineração, exploração madeireira ilegal, dentre outros, é possível modelar o efeito de causas indiretas que afetam essas áreas, como a pressão demográfica, políticas públicas ineficazes, ou recursos econômicos mal direcionados (Geist e Lambin, 2002).

Em situações em que o fator de estudo se distribui espacialmente, a adição do termo de fragilidade, contudo, não é suficiente para capturar a relação com efeitos aleatórios. Essa relação está intimamente ligada à dependência espacial dos dados, ou seja, a estrutura espacial carrega a informação de efeitos aleatórios que estão relacionados ao fator de estudo. Nesse cenário, são aplicáveis os modelos de

fragilidade espacial, e, para essas situações, a distribuição de probabilidade deve ser substituída por um modelo que incorpora a dependência espacial dos dados. Nestes modelos, são assumidos, por construção, a distribuição normal para o termo de fragilidade. Banerjee (2003) apresenta alguns modelos autorregressivos que foram criados com esse intuito, como os modelos CAR, ICAR e SAR.

Em estudos acerca do monitoramento de atividades humanas e seus efeitos no meio ambiente, os modelos de fragilidade podem ser utilizados para determinar o efeito de covariáveis econômicas e sociais no desmatamento, um problema constantemente estudado e monitorado (Fearnside, 2005; Saha et al., 2020). Nesse contexto, a aplicação de modelos de fragilidade se apresenta como uma contribuição valiosa, pois permite uma análise aprofundada do impacto de covariáveis sociais e econômicas no desmatamento, além da análise de efeitos relacionados a fatores aleatórios.

Sob esse cenário de análise e contextualização do desmatamento, este trabalho foi dividido em dois capítulos, sendo o primeiro referente a uma revisão sobre o desmatamento e suas dinâmicas, juntamente com técnicas de análise de sobrevivência e modelagem espacial. No segundo capítulo, é utilizado o modelo de fragilidade, com o objetivo de avaliar o efeito de covariáveis socioeconômicas e ambientais no tempo até a ocorrência do desmatamento para a região sul do estado do Amazonas.

Referências

BALAN, T. A.; PUTTER, H. A tutorial on frailty models. **Statistical Methods in Medical Research**, v. 29, n. 11, p. 3424–3454, nov. 2020.

BANERJEE, S. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. **Biostatistics**, v. 4, n. 1, p. 123–142, 1 jan. 2003.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. 2. ed. São Paulo, SP: Editora Edgard Blucher, 2024.

DIAS, F.; SUHADOLNIK, N.; CAMARGO, H.; DA SILVA, S. Predicting the pulse of the Amazon: Machine learning insights into deforestation dynamics. **Journal of Environmental Management**, v. 362, p. 121359, jun. 2024.

DOMINGUEZ, D.; DEL VILLAR, L. D.; PANTOJA, O.; GONZALEZ, M. Forecasting Amazon Rain-Forest Deforestation Using a Hybrid Machine Learning Model. **Sustainability**, v. 14, n. 2, p. 691, 9 jan. 2022.

FEARNSIDE, P. M. Deforestation in Brazilian Amazonia: History, Rates, and Consequences. **Conservation Biology**, v. 19, n. 3, p. 680–688, jun. 2005.

GEIST, H. J.; LAMBIN, E. F. Proximate Causes and Underlying Driving Forces of Tropical Deforestation. **BioScience**, v. 52, n. 2, p. 143, 2002.

HARGRAVE, J.; KIS-KATOS, K. Economic Causes of Deforestation in the Brazilian Amazon: A Panel Data Analysis for the 2000s. **Environmental and Resource Economics**, v. 54, n. 4, p. 471–494, abr. 2013.

HOUGAARD, P. Frailty models for survival data. **Lifetime Data Analysis**, v. 1, n. 3, p. 255–273, 1995.

JUSYS, T. Fundamental causes and spatial heterogeneity of deforestation in Legal Amazon. **Applied Geography**, v. 75, p. 188–199, out. 2016.

KYERE-BOATENG, R.; MAREK, M. V. Analysis of the Social-Ecological Causes of Deforestation and Forest Degradation in Ghana: Application of the DPSIR Framework. **Forests**, v. 12, n. 4, p. 409, 29 mar. 2021.

MONTEIRO, A. M. V.; RENNO, C. D.; ALMEIDA, C. A.; VALERIANO, D. M.; SOARES, J. V.; MAURANO, L. E. P.; ESCADA, M. I. S.; AMARAL, S.; PINHEIRO, T. F. **Relatório Técnico-Científico**. 20 maio 2008.

SAHA, S.; SAHA, M.; MUKHERJEE, K.; ARABAMERI, A.; NGO, P. T. T.; PAUL, G. C. Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: A case study at the Gumani River Basin, India. **Science of The Total Environment**, v. 730, p. 139197, ago. 2020.

CAPÍTULO 1: REFERENCIAL TEÓRICO

Neste capítulo, será apresentada uma introdução e contextualização do desmatamento, explorando suas dinâmicas, consequências e estratégias de preservação. Também será apresentada uma revisão bibliográfica sobre análise de sobrevivência e a abordagem bayesiana com ênfase na aplicação dos modelos de fragilidade.

1.1. O desmatamento e a biodiversidade

Nas últimas décadas, o desmatamento tem causado cada vez mais pressões em grandes ecossistemas na Amazônia. Conforme surgiram necessidades de intervenções para se conduzir ações de preservação contra o desmatamento no Brasil, foi criado, em 1988, o sistema PRODES (Assunção et al., 2023; Monteiro et al., 2008) conduzido pelo INPE - Instituto Brasileiro de Pesquisas Espaciais. Esse projeto foi criado com o intuito de reportar informações anuais sobre o desmatamento com o objetivo de levantar dados para que fosse possível o monitoramento e controle das áreas florestais (Monteiro et al., 2008).

Com o passar dos anos, no entanto, informações anuais não estavam se mostrando eficientes em ajudar nas ações de controle rápido do desmatamento. O projeto DETER (Assunção et al., 2023; Diniz, 2015) surge então com uma nova necessidade de se obter informações mais rapidamente com relação às ocorrências de desmatamento. Criado em 2004, o projeto DETER consegue capturar qualquer intervenção na floresta superior a 25 hectares. A partir de 2017, o INPE passou a operar com uma nova versão do DETER, conhecido como DETER-b (Assunção et al., 2023; Diniz, 2015). Com os levantamentos dos projetos DETER e DETER-b foi possível monitorar ocorrências de desmatamento quase que em tempo real. O objetivo para a criação e utilização desses projetos é ter um sistema de alerta que reporta informações rapidamente a órgãos de preservação como o IBAMA.

O Terraclass (Neves et al., 2020), é outro projeto criado e monitorado pelo INPE que iniciou em 2010. O seu surgimento está associado à necessidade de qualificação do desmatamento observado. Para maior controle de informações, era necessário ter uma visão mais detalhada não só do desmatamento, mas de fatores que contribuíam para esse aumento observado na área desmatada. O projeto Terraclass concentra uma grande variedade de informações e mapas interativos que permite uma boa

visualização de áreas desmatadas, em conjunto à classificação do território quanto a fatores de interesse no controle do desmatamento, como mineração, agricultura e população.

Devido a diversidade de fatores que influenciam o desmatamento, a análise e monitoramento de ocorrências são frequentemente alvos de estudos que têm por objetivo descrever as regiões relacionadas às ocorrências observadas, para que sejam direcionadas e melhoradas as políticas de proteção e conservação ambiental (Barona et al., 2010; Mena et al., 2006).

Para que maiores esforços sejam direcionados à proteção e à preservação, são discutidas parcerias com produtores, empresas privadas e ações não governamentais, para atuar paralelamente a ações e atuais leis do governo. O projeto Responsible Soy Project (Jung e Polasky, 2018; Panwar et al., 2023) é uma iniciativa desse esforço, que atua no combate ao desmatamento por ter caráter ambiental, social e economicamente sustentável.

O controle do avanço do desmatamento em grandes regiões florestais, pode também influenciar diretamente na diversidade da fauna local, uma vez que animais silvestres, principalmente espécies ameaçadas de extinção, dependem de habitats específicos, e muitas vezes estão presentes em apenas dois ou três municípios. Alguns trabalhos que possuem esse direcionamento, propõem avaliar o desmatamento com relação a um único fator, e aprofundar um pouco mais com relação a essa associação. Sonter et al. (2017), propõem avaliar o desmatamento na região da Amazônia com relação às operações de mineração, e como essa dinâmica se comporta ao longo do tempo.

Santos et al. (2021) propuseram avaliar o desmatamento por meio da área anual desmatada em relação a covariáveis socioeconômicas, como o número de pessoas residindo na zona rural, padronizado pela área de cada município. A análise é realizada considerando 550 municípios localizados no bioma Amazônico com relação a covariáveis socioeconômicas, ecológicas e físicas. Para essa avaliação, foi utilizado o modelo de Regressão Geograficamente Ponderada (GWR), uma extensão do modelo de regressão linear, que emprega uma matriz de pesos W derivada de uma função kernel, para ponderar os coeficientes em relação a cada município. Dessa forma, os coeficientes de regressão estimados variam conforme a localização geográfica. Neste trabalho, os autores identificaram a presença de um grupo formado

por áreas desmatadas na porção sul da Amazônia, região conhecida como o arco do desmatamento.

Jusys (2016) utiliza também a Regressão Geograficamente Ponderada (GWR), para avaliação de covariáveis socioeconômicas, ambientais e avaliação da distribuição espacial do desmatamento na Amazônia Legal. Com esse mesmo propósito, Monjardin-Armenta et al. (2020) buscam avaliar o desmatamento por meio de uma análise espacial, utilizando técnicas de avaliação multicritério, em que o objetivo foi considerar o comportamento futuro do desmatamento e reflorestamento e realizar o seu mapeamento.

Sob essa perspectiva, é possível perceber que fatores sociais e econômicos podem interagir entre si e produzir um efeito conjunto a partir dessa interação. Silva-Junior et al. (2023) descrevem a importância de terras indígenas para o controle do desmatamento e a implicação desse efeito nessas áreas. Além disso, é discutido como outros fatores podem causar pressões e comprometer as terras indígenas, afetando diretamente o desmatamento, como as atividades de mineração ilegal.

O monitoramento do desmatamento por satélites muitas vezes pode ser comprometido devido à cobertura por nuvens. Sales et al. (2022) propõem a utilização do modelo de Cox (Cox, 1972) para a avaliação de como a cobertura por nuvens afeta o monitoramento do desmatamento por satélites na Amazônia Legal. Esses autores fazem menção que, nos dias atuais, esse impacto é reduzido devido ao maior número de satélites que fazem esse monitoramento em órbita terrestre.

Dados da Global Forest Watch – GFW (Mulissa et al., 2024) mostram que, entre os anos de 2001 a 2023, foi registrado um total de 488 milhões de hectares de perda na cobertura arbórea em nível global, o que resultou em uma redução de 12% na cobertura arbórea em comparação ao ano 2000. É evidente que outras causas podem ser citadas, como queimadas, que provocam o mesmo efeito. Segundo dados da própria GFW, essa perda corresponde ao equivalente à emissão de aproximadamente 207 gigatoneladas de dióxido de carbono (CO₂) na atmosfera (Kinnebrew et al., 2022; Mulissa et al., 2024).

Com todos os esforços e estudos já realizados, é notória a importância do estudo do desmatamento, tendo em vista toda a mudança de comportamento, contextos políticos, sociais e socioeconômicos ao longo dos anos.

1.2. A análise de sobrevivência

A análise de sobrevivência é um conjunto de técnicas estatísticas que buscam estudar o comportamento do tempo até a ocorrência de algum evento de interesse. Em geral, o grande objetivo é fazer essa verificação na presença de covariáveis de interesse. Outro fator característico da análise de sobrevivência são as censuras, em que indivíduos que não experimentaram ou que por algum motivo foram impedidos de experimentar o evento de interesse são denominados observações censuradas (Colosimo e Giolo, 2024). O mecanismo de censura pode ser representado por duas variáveis aleatórias independentes $(T_i; C_i)$, isto é, a distribuição de C não carrega informação sobre T , sendo T o tempo até o evento e C o tempo até a censura. Desta forma, o que se observa para cada indivíduo é $t = \min(T_i; C_i)$ e δ é dado por:

$$\delta_i = \begin{cases} 1; & T_i \leq C_i \\ 0; & T_i > C_i \end{cases}$$

Desse modo, alguns tipos de censura podem ser observados, em que a observação pode não experimentar o evento de interesse de maneiras diferentes. São denominadas censuras à esquerda, as observações em que o evento de interesse já havia ocorrido quando a unidade experimental foi observada. São denominadas censuras à direita, quando o tempo de falha acontece após o estudo acabar, ou nem mesmo acontece. As censuras do tipo intervalar acontecem quando se observa um intervalo em que se sabe que a observação aconteceu, mas não se sabe o tempo exato dessa ocorrência (Colosimo e Giolo, 2024).

A função densidade de probabilidade $f(t)$ pode ser definida como a probabilidade de uma observação sofrer o evento de interesse dado um intervalo imediato de tempo. Sendo T uma variável aleatória contínua, positiva, e $f(t)$ a sua função densidade de probabilidade:

$$f(t) = \lim_{\epsilon \rightarrow 0^+} \frac{\Pr(t \leq T < t + \epsilon)}{\epsilon}$$

A função de distribuição acumulada $F(t)$ da variável aleatória T é dada pela probabilidade de o evento ocorrer até o tempo t , sendo:

$$F(t) = \Pr(T \leq t)$$

A função de sobrevivência $S(t)$ é a probabilidade de uma observação não falhar por no mínimo um tempo t estabelecido previamente no estudo, em que:

$$S(t) = P(T > t)$$

As funções de sobrevivência e de distribuição acumulada estão relacionadas por meio das seguintes relações (Colosimo e Giolo, 2024):

$$S(t) = \Pr(T > t) = 1 - \Pr(T \leq t) = 1 - F(t)$$

A função taxa de falha mede a taxa instantânea de falha em determinado tempo t , dado que o evento de interesse ainda não ocorreu. A função taxa de falha relaciona-se às demais funções por meio da seguinte expressão:

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}(\log S(t))$$

em que, a função taxa de falha acumulada pode ser obtida por:

$$\Lambda(t) = \int_0^t \lambda(u) du = -\log S(t)$$

1.2.1. O estimador não paramétrico de Kaplan-Meier

A análise descritiva dos dados é amplamente utilizada e fortemente recomendada para a verificação inicial do comportamento dos dados. Essa verificação permite, de antemão, observar padrões e valores discrepantes, comportamentos não comuns, bem como prever com maior precisão técnicas que podem ser aplicadas posteriormente para analisar os dados. Apesar disso, em problemas em que há a presença de elementos censurados, essa análise é dificultada, por exemplo na criação de um histograma. Nesse caso, o eixo x é comprometido pela presença de censuras, pois o seu formato não permite a captura dos dados presentes nas censuras.

O estimador de Kaplan-Meier é uma técnica estatística muito utilizada que permite incorporar a informação das falhas e censuras e não prevê a utilização de uma distribuição de probabilidades, sendo, portanto, uma técnica não paramétrica. Proposto por Kaplan e Meier (1958), a partir desse estimador é possível obter a probabilidade de sobrevivência estimada para cada tempo t , com relação ao número de elementos censurados (Colosimo e Giolo, 2024).

Seja q_j a probabilidade de falha no intervalo $(t_{j-1}, t_j]$ para $j = 1, 2, \dots, k$. Logo:

$$q_j = P(T \in (t_{j-1}, t_j] | T > t_{j-1})$$

$$\hat{q}_j = \frac{\text{total de falhas em } (t_{j-1}, t_j]}{\text{total de indivíduos sob risco em } t_{j-1}}$$

A fórmula geral para a sobrevivência estimada pelo estimador de Kaplan-Meier é, portanto:

$$\hat{S}(t) = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

em que n_j representa o total de indivíduos sob risco, ou seja, os que não falharam e não foram censurados até t_{j-1} , e d_j representa o total de falhas em $(t_{j-1}, t_j]$.

No estimador de Kaplan-Meier não é possível modelar covariáveis como em um modelo de regressão, e a estimação da sobrevivência é feita para cada covariável separadamente. Com isso, é possível estimar a probabilidade de sobrevivência em cada tempo observado considerando os elementos censurados. Outros estimadores não paramétricos podem ser encontrados na literatura, como o estimador de Nelson-Aalen (Colosimo e Giolo, 2024).

1.2.2. Técnicas paramétricas para análise de sobrevivência

As técnicas paramétricas para analisar dados de sobrevivência são de grande importância na presença de covariáveis de interesse, uma vez que, na maioria dos casos, o intuito é avaliar o efeito de cada uma delas na variável resposta. Além disso, é possível estimar a sobrevivência em cada ponto do tempo analisado, diferentemente

de uma técnica não paramétrica que realiza essa estimação para intervalos de tempo, com relação ao número de indivíduos totais e o número de indivíduos sob risco naquele momento (Colosimo e Giolo, 2024).

Diversos modelos foram propostos para modelar o tempo até a ocorrência de um evento de interesse (Carroll, 2003). É importante observar que a distribuição utilizada para essa finalidade deve ter suporte positivo, uma vez que o tempo é medido por valores positivos, e ser contínua em todo o seu domínio. Dentre os modelos mais utilizados, se destacam os modelos Exponencial, Weibull, Log-normal e Log-logístico.

A tabela abaixo representa as funções de densidade de probabilidade e funções de sobrevivência para esses modelos:

Tabela 1 - Principais modelos paramétricos de sobrevivência com identificação de sua função de densidade de probabilidade e função de sobrevivência associada.

Modelos	Função densidade de probabilidade	Função de sobrevivência
Exponencial $\lambda > 0, t \geq 0$	$f(t) = \lambda \exp\{-\lambda t\}$	$S(t x) = \exp\left\{-\frac{t}{\exp(\beta_0 + \boldsymbol{\beta}^T \mathbf{x})}\right\}$
Weibull $\gamma, \lambda > 0, t \geq 0$	$f(t) = \lambda \gamma t^{\gamma-1} \exp\{-\lambda t^\gamma\}$	$S(t x) = \exp\left\{-\left(\frac{t}{\exp(\beta_0 + \boldsymbol{\beta}^T \mathbf{x})}\right)^{\frac{1}{\sigma}}\right\}$
Log-normal $\sigma > 0, t \geq 0$	$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{\ln(t) - \mu}{\sigma}\right)^2\right\}$	$S(t x) = \Phi\left(\frac{-\ln(t) + \beta_0 + \boldsymbol{\beta}^T \mathbf{x}}{\sigma}\right)$
Log-logística $\gamma, \lambda > 0, t \geq 0$	$f(t) = \frac{\gamma t^{\gamma-1} \lambda}{(1 + \lambda t^\gamma)^2}$	$S(t x) = \frac{1}{1 + \left(\frac{t}{\exp(\beta_0 + \boldsymbol{\beta}^T \mathbf{x})}\right)^{\frac{1}{\sigma}}}$

Φ = função de distribuição acumulada da distribuição normal padrão; ln = Logaritmo natural.

Fonte: (Martins Filho et al., 2023).

A seleção do melhor modelo paramétrico é realizada com base em estatísticas como o AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion). A qualidade do ajuste pode ser feita pelos resíduos de Cox-Snell, em que a distribuição dos resíduos deve se aproximar de uma distribuição exponencial padrão.

1.3. O modelo ICAR

Propostos por Besag (1974), os modelos autorregressivos dividem-se em algumas classes, a depender de especificações sutis em sua utilização. O modelo ICAR (*Intrinsic Conditional Autorregressive Model*) foi introduzido como uma classe de modelos autorregressivos, e possui o objetivo de inserir uma estrutura de dependência espacial para as covariáveis não observadas que representam o termo de efeito aleatório. Estes modelos passaram a ser popularmente conhecidos a partir do surgimento dos métodos MCMC (*Monte Carlo Markov Chain*), a partir da década de 1990, e que utilizam as especificações de distribuições *a priori* para os hiperparâmetros do modelo. O atraso ocorrido para a popularização dos modelos autorregressivos condicionais está ligado à complexidade e a dependência dos recursos computacionais para a sua utilização (Assunção e Krainski, 2009; Lavine e Hodges, 2012).

Este modelo é especificado considerando uma matriz \mathbf{A} de dimensão $L \times L$, em que L é o número total de áreas formadas com características semelhantes entre si. O número de vizinhos do modelo ICAR é uma informação capturada diretamente da matriz de vizinhanças. O objetivo é inserir a dependência espacial com relação à uma matriz binária de vizinhanças, em que recebe 1 se a área i possuir fronteira com a área j , e 0 caso contrário (Banerjee, 2003). Essa relação pode ser expressa por:

$$(w_l | \mathbf{w}_{(-l)}) \sim N\left(\bar{w}_l; \frac{1}{(\phi n_l)}\right), \quad l = 1, \dots, L$$

em que w_l (valor associado à correlação da estrutura das áreas no mapa e as covariáveis não observadas) está ligado à l -ésima área do mapa, n_l representa o número de vizinhos associados à área em questão, e ϕ representa o parâmetro de precisão do modelo.

1.4. Matriz de vizinhanças

A análise de dados espaciais possui grande flexibilidade quanto a incorporação da variabilidade dos dados em um modelo. Uma delas é a análise pontual, em que cada indivíduo é analisado separadamente dos demais. Outra, é a análise de dados de áreas, que está geralmente associada a grupos de indivíduos em comum, em que

são considerados polígonos ou porções de áreas contendo as concentrações de cada grupo (Máquina et al., 2019).

O objetivo em estudar grupos de indivíduos como alternativa a análise caso a caso, é a associação esperada entre grupos com relação aos tempos de sobrevivência. Além disso, a interpretação é relevante uma vez que se relaciona à variável resposta. Neste caso, a matriz de vizinhanças, ou matriz de adjacências, corresponde ao desenho da área representada pelos polígonos de estudo, porém de forma matricial. O nome das linhas e colunas da matriz é representado pela nomenclatura de cada polígono. Sendo assim, se o espaço total da área conter 12 polígonos, a matriz de vizinhanças resultante terá dimensão 12 x 12. Esta matriz é formada por 0 e 1, sendo 0 caso não haja relação de vizinhança de um polígono com o outro, ou 1 caso essa relação de vizinhança seja identificada.

O entendimento do conceito de vizinhança pode ser determinado de várias formas, e muitas vezes deve ser estabelecido pelo pesquisador. No entanto, existem maneiras comuns de se estabelecer o conceito de vizinhança. O caso conhecido como *queen's case* considera como vizinho todo polígono que tiver ao menos um ponto de fronteira, enquanto o *rook's case* considera como vizinho todo polígono que possuir mais que um ponto de fronteira (Bivand, et al., 2015). Isso significa que, ao converter um arquivo de polígonos para uma matriz de vizinhanças, o desenho da área formada pelo polígono é perdido, porém é mantida a estrutura espacial dos dados.

1.5. O modelo de fragilidade compartilhada

Os modelos de análise de sobrevivência clássica assumem a independência dos tempos observados e também que as censuras são não informativas. A não informação indica que o fato de o evento não ter ocorrido até determinado momento, não adiciona informações sobre o risco desse evento ocorrer.

Introduzidos por Vaupel et al. (1979), os modelos de fragilidade buscam adicionar no modelo a heterogeneidade não observada nos dados. Em outras palavras, o interesse é inserir a informação de covariáveis latentes com base em algum outro termo correlacionado.

O modelo de Cox (Cox, 1972) é uma abordagem bastante utilizada em análise de sobrevivência, em que é utilizado para quantificar o efeito das covariáveis no tempo de sobrevivência da variável de estudo (Colosimo e Giolo, 2024). Para essa

abordagem, a pressuposição de riscos proporcionais para as covariáveis deve ser satisfeita.

Um exemplo de aplicação do modelo de fragilidade compartilhada são os modelos de riscos proporcionais. Nesse cenário, o modelo de fragilidade compartilhada pode ser considerado uma extensão do modelo de Cox, em que um termo de efeito aleatório e multiplicativo é adicionado ao termo geral (Balan e Putter, 2020). Outra aplicação para os modelos de fragilidade está relacionada aos modelos de tempo de vida acelerados (AFT). Sob essa perspectiva, um modelo paramétrico de fragilidade Weibull (AFT), pode ser definido da seguinte maneira (Pereira et al., 2024):

$$f(t) = \alpha t^{\alpha-1} \lambda \exp(-\lambda t^\alpha)$$

em que $\alpha > 0$ e $\lambda > 0$. Desse modo, $\lambda = \exp(\eta)$ e $\eta = \mathbf{x}'_{li} \boldsymbol{\beta} + w_l$, em que w_l representa o termo que caracteriza o modelo de fragilidade (Rue et al., 2009). Nesse sentido, α representa o parâmetro de forma da distribuição, λ representa o parâmetro de escala da distribuição e η representa o preditor linear. Além disso, os coeficientes presentes em $\boldsymbol{\beta}$ associam-se às covariáveis, enquanto os termos w_l representam os termos de efeito aleatório associados ao modelo de fragilidade.

Várias distribuições de probabilidade podem ser empregadas para modelar o termo de efeito aleatório. Dentre essas, a distribuição gama se destaca por sua flexibilidade e facilidade para assumir diferentes formas. No entanto, no contexto da modelagem da fragilidade espacial por meio do modelo ICAR, a distribuição normal é assumida, por definição, para os termos de efeito aleatório.

O intuito é modelar a variabilidade latente dos eventos observados em adição às covariáveis identificadas. Desse modo, quanto mais "frágil" for uma determinada área, maiores serão as chances de se obter tempos de sobrevivência menores, considerando a área total. Nos modelos de riscos proporcionais, Balan e Putter (2020) alertam que, ao utilizar covariáveis não observadas representadas pelas fragilidades, as covariáveis latentes de fragilidade e a pressuposição de riscos proporcionais podem estar confundidas. Isso porque a inserção do termo aleatório pode acomodar efeitos que não são proporcionais ao longo do tempo, e, essa avaliação pode se tornar ainda mais desafiadora quando se trabalha com grupos pequenos.

A interpretação dos coeficientes das covariáveis inseridas no modelo, está diretamente ligada ao valor do termo de fragilidade daquele grupo, uma vez que

grupos com valores idênticos para as covariáveis, podem possuir diferentes probabilidades de sobrevivência, devido a atuação do termo aleatório (Colosimo e Giolo, 2024; Govindarajulu et al., 2011). Essa interpretação também pode ser feita com relação ao risco, a depender dos critérios adotados na análise e das especificações do modelo.

Os valores correspondentes às fragilidades em cada área são ditos como relativos e não absolutos (Colosimo e Giolo, 2024). Isso se dá porque a interpretação desse valor para uma área depende das demais, desta forma, esse valor não pode ser interpretado de forma absoluta. Nesse sentido, o seu significado está sujeito às comparações de diferenças com as demais regiões.

1.5.1. Uma abordagem bayesiana

É comum a função de verossimilhança envolver expressões complexas de se trabalhar algebricamente. Isto porque, é necessário englobar toda a estrutura criada para representar os grupos de indivíduos semelhantes, e, ao mesmo tempo, associar às covariáveis observadas e os termos de efeito aleatório. Nesses casos a abordagem bayesiana para estimação dos parâmetros é usada com maior frequência, visto que é mais flexível com relação à abordagem clássica, em grande parte dos casos, para modelos complexos. Uma desvantagem que deve ser levada em consideração é que a abordagem bayesiana pode ser dificultada em muitos contextos, por exigir um grande custo computacional (Lindgren e Rue, 2015).

A inferência bayesiana consiste em utilizar uma abordagem baseada no teorema de Bayes, para se realizar inferências sobre os parâmetros do modelo. A abordagem paramétrica de análise de dados envolve a especificação completa da distribuição do modelo de probabilidades associado aos dados $p(y|\theta)$, $\theta \in \Theta$ (Hanagal e Dabade, 2013; Pan, 2001; Umeta et al., 2022). No contexto de modelos de fragilidade espacial, a verossimilhança pode ser descrita como:

$$L(\mu, \sigma) = \prod_{l=1}^L \prod_{i=1}^{n_l} [f(t_{l,i}|x_{li})]^{\delta_{li}} S(t_{l,i}|x_{li})^{1-\delta_{li}}$$

em que δ_{li} representa o indicador de falha ou censura, em um cenário formado por L grupos e n_l indivíduos em cada grupo. Seja ψ os hiperparâmetros do modelo

associados à distribuição *a priori* $p(\theta)$. Dessa maneira, a construção do modelo pode ser expressa como:

$$p(y) = \int_{\theta} L(y|\theta)p(\theta)d\theta$$

Entretanto,

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

Assim,

$$p(\theta) = \int_{\psi} p(\theta|\psi)p(\psi)d\psi$$

em que $p(\theta|y)$ representa a distribuição *a posteriori* condicionada aos hiperparâmetros do modelo (Spiegelhalter et al., 2002).

Os métodos MCMC (*Monte Carlo Markov Chain*) têm sido amplamente utilizados como uma boa alternativa para se trabalhar com inferência na abordagem bayesiana, uma vez que permitem simular amostras da distribuição de interesse e a partir delas verificar o comportamento de convergência dos parâmetros (Alvares et al., 2021). A partir de cada simulação gerada, são guardados os valores dos parâmetros obtidos de modo que, ao final do processo, cada parâmetro possui uma cadeia de valores associada a ele. Para a realização do processo de simulação de amostras, algumas opções podem ser os softwares OpenBugs (Carroll et al., 2015; Cowles, 2013) ou WinBugs (Ntzoufras, 2009), em que ambos possuem fácil integração com o software R.

Outra metodologia amplamente utilizada é o método INLA (*Integrated Nested Laplace Approximation*). Desenvolvido por Rue et al. (2009), já é amplamente difundido e consolidado na literatura para se lidar com a função de verossimilhança para obtenção da distribuição *a posteriori* de interesse em diferentes modelos (Martino e Rue, 2010). Dentre estes, estão os modelos gaussianos latentes, que incorporam covariáveis não observadas ao modelo, e assumem distribuição normal como distribuição *a priori*. Ao contrário dos métodos MCMC, a obtenção da distribuição *a posteriori* é feita utilizando a aproximação numérica de Laplace. Maiores detalhes, definições e aplicações do método podem ser encontrados em Lindgren e Rue (2015), Martins et al. (2013) e Moraga et al.(2021).

A interpretação dos coeficientes do modelo de fragilidade não deve ser realizada diretamente como no modelo paramétrico, uma vez que a parametrização utilizada pelo pacote *INLA* é diferente para os modelos de tempo de vida acelerados. Sejam os valores ζ os valores estimados pelo modelo associados ao efeito das covariáveis, a qual deseja-se transformar para uma interpretação com relação a aceleração ou desaceleração do tempo de sobrevivência (Wang et al., 2018). Sejam também os valores β 's referentes aos coeficientes estimados pelo modelo paramétrico com esse mesmo propósito. A relação entre ambos os coeficientes estimados é descrita pela seguinte equação :

$$\zeta = \frac{-\beta}{\sigma}$$

Nesse sentido, a interpretação dos coeficientes calculados em ambas as abordagens deve ser feita de forma cautelosa de maneira a seguir essa relação. Os modelos paramétricos estimados pelo pacote *survival* (Therneau, 2001), podem ser obtidos por meio do pacote *INLA*, observando-se as regras para interpretação dos coeficientes, em que, β representa os coeficientes estimados pela função *survreg*, e ζ representam os coeficientes estimados pelo *INLA*.

1.5.2. Deviance Information Criterion – DIC

O DIC (Spiegelhalter et al., 2002) é considerado uma medida alternativa bayesiana ao AIC (Akaike Information Criterion) para comparar modelos e verificar qual se adequa melhor aos dados (Gelman et al., 2019). Isso porque a definição e utilização de ambos são bastante semelhantes. Quanto melhor o modelo se ajustar aos dados, menor será o valor do DIC. O seu cálculo é baseado na *deviance* e no número efetivo de parâmetros p_D . O objetivo é considerar o ajuste aos dados e também penalizar modelos mais complexos.

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

em que $\theta = (\alpha, \beta, \phi)$ representa os parâmetros do modelo. $\bar{D}(\theta)$ representa a *deviance* média *a posteriori*. O número efetivo de parâmetros do modelo pode ser expresso por:

$$p_D = \overline{D(\theta)} - D(\bar{\theta})$$

sendo,

$$D(\theta) = -2\log[p(y|\theta)] + 2\log[f(y)]$$

em que $E(Y) = \mu(\theta)$ e $f(y) = p[y|\mu(\theta) = y]$.

1.5.3. O fator de Bayes

O fator de Bayes é uma métrica utilizada na inferência bayesiana para comparar dois modelos, e verificar qual se ajusta melhor aos dados. É baseado no teorema de Bayes e no cálculo da verossimilhança marginal *a posteriori* para ambos os modelos (Schmalz et al., 2023). De acordo com Hoijtink (2019), considerando dois modelos candidatos, o objetivo é quantificar o quão próximos estão os dados do modelo 1, ou do modelo 2.

Para compor o fator de Bayes, é utilizada a razão entre duas probabilidades condicionais *a posteriori*, logo:

$$BF = \frac{p(\mathbf{y}|\psi_1)}{p(\mathbf{y}|\psi_2)} = \frac{\int_{\theta_1} p(\mathbf{y}|\theta_1)p(\theta_1|\psi_1)d\theta_1}{\int_{\theta_2} p(\mathbf{y}|\theta_2)p(\theta_2|\psi_2)d\theta_2} = \frac{\frac{p(\psi_1|\mathbf{y})p(\mathbf{y})}{p(\psi_1)}}{\frac{p(\psi_2|\mathbf{y})p(\mathbf{y})}{p(\psi_2)}} = \frac{p(\psi_1|\mathbf{y})p(\psi_2)}{p(\psi_2|\mathbf{y})p(\psi_1)}$$

em que, θ_1 corresponde aos parâmetros do modelo 1 e ψ_1 são os hiperparâmetros do modelo 1, e igualmente para o modelo 2.

É evidente que, a razão proposta pelo fator de Bayes corrobora uma análise na qual, se não houver diferença significativa entre os modelos, estará próxima de 1. Variações discrepantes, tanto para mais quanto para menos em relação a esse valor, indicam uma maior suspeita de que um dos modelos é melhor. No entanto, não existe uma única abordagem com relação às categorias de valores para a interpretação do fator de Bayes. Sendo assim, podem ser encontradas diferentes disposições sobre essa medida em diversos contextos. Kass e Raftery (1993) apresentam uma tabela comumente utilizada para a interpretação do fator de Bayes:

Tabela 2 – Tabela para identificação e interpretação do fator de Bayes.

log (BF)	BF	Evidência
0 a 0,5	1 a 3,2	Não significativa
0,5 a 1	3,2 a 10	Substancial
1 a 2	10 a 100	Forte
> 2	> 100	Decisiva

Fonte: (Kass e Raftery, 1993).

Considerando tal perspectiva, a magnitude do valor encontrado é interpretada e a diferença entre os modelos é conduzida a um cenário em que será escolhido um deles para a utilização, com relação à sua adequação aos dados.

1.5.4. O método K-Means

O método K-Means é um algoritmo de classificação amplamente utilizado e difundido. O seu principal objetivo é encontrar e classificar os indivíduos conforme a distância euclidiana para o centroide do grupo (Likas et al., 2003; Vhudzijena, 2023). Inicialmente é escolhida uma sequência de agrupamento que minimiza a soma dos quadrados dentro de cada grupo.

$$SSE = \sum_{i=1}^n \|x_i - c_{k(i)}\|^2$$

Nesse sentido, n representa o número total de pontos, x_i é o i -ésimo ponto no espaço de estudo e $c_{k(i)}$ representa o centroide do grupo a qual o ponto x_i pertence. A partir do número ótimo de grupos k , são escolhidos os centroides iniciais de maneira aleatória. Os agrupamentos iniciais são calculados atribuindo os pontos ao centroide mais próximo. No próximo passo, os centroides precisam ser recalculados como a média daquele grupo atual. Novamente o processo iterativo é repetido até que a soma de quadrados não se altere significativamente e o método atinja a convergência.

Colosimo e Giolo (2024) explicam que, em ambos os casos univariado e multivariado, existe uma variável grupos que, por construção do modelo, é a variável que representa o motivo pelo qual os tempos são correlacionados, seja por

dependência espacial, fatores genéticos, entre outros. Para o caso multivariado, que é de interesse nesse estudo, os grupos podem ser representados como uma covariável, porém no ajuste do modelo são tratados de forma peculiar para englobar os efeitos das fragilidades. No entanto, em alguns casos o pesquisador deseja inserir a heterogeneidade não observada nos dados ao modelo, mas não tem uma ideia inicial dos grupos a serem usados no modelo de fragilidade. Govindarajulu e Bedi (2022) apresentam uma abordagem de aplicação do método K-Means nos modelos de fragilidade para a obtenção desses grupos, e o método demonstrou ser uma alternativa para esse caso específico. Nesse sentido, os grupos são gerados a partir do método K-Means e usados posteriormente no modelo de fragilidade.

Uma medida de avaliação do número de grupos, denominada silhueta, é muito utilizada no método K-Means, indicando o quanto um ponto se aproxima do seu próprio grupo com relação aos demais. Essa métrica pode ser obtida pela seguinte relação:

$$s(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]}$$

em que $s(i)$ representa a silhueta média observada para o ponto i . O termo $a(i)$, portanto, representa a média das distâncias do ponto i para todos os demais pontos dentro do mesmo grupo. Já o termo $b(i)$ representa a média das distâncias do ponto i para todos os pontos do cluster mais próximo. O objetivo com a aplicação do método K-Means é que a informação aleatória presente na configuração espacial do desmatamento seja incorporada ao modelo com a adição do termo de fragilidade.

1.5.5. Seleção de variáveis

O método de Collett para seleção de covariáveis (Collett, 1994) consiste na condução de passos para verificar a importância de cada variável para as análises elencadas previamente. O objetivo é otimizar o processo de construção do modelo, para que sejam mantidas somente as covariáveis que realmente contribuem para o ajuste global. Nesse método, a critério do pesquisador, uma variável pode ser mantida no modelo mesmo que ela não seja significativa, caso haja uma justificativa teórica forte para isso.

A seleção do melhor modelo pode ser feita a partir do teste da razão de verossimilhanças, considerando modelos encaixados. Quando não for esta a situação, pode ser utilizado o AIC e/ou o BIC. No caso da abordagem bayesiana, o DIC e/ou o fator de Bayes.

Considerando uma adaptação do método de Collett (Collett, 1994), os seguintes passos podem ser destacados para a seleção das covariáveis:

- **Passo 1:** Ajuste inicial contendo cada variável separadamente das demais. Nesse passo cada variável deve estar contida em um ajuste individual. Cada ajuste individual deve ser comparado com o modelo nulo, sem covariável. As covariáveis que se mostrarem melhorar o modelo com relação ao modelo nulo são selecionadas nesse passo.
- **Passo 2:** A partir de todas as covariáveis no passo anterior, é feito um ajuste com todas as covariáveis selecionadas. A partir desse ponto, são retiradas as covariáveis uma a uma e os modelos comparados com o modelo completo. Serão selecionadas as covariáveis que mostrarem contribuição global com o modelo na presença das demais covariáveis.
- **Passo 3:** É feito um ajuste com todas as covariáveis selecionadas no passo 2. A partir desse ponto, são adicionadas as covariáveis descartadas no passo 1, com o intuito de verificar se elas contribuem significativamente para o ajuste global do modelo. Serão adicionadas as covariáveis descartadas no passo 1, e que se mostrarem significativas no ajuste global.
- **Passo 4:** A partir de todas as covariáveis selecionadas no passo 3, são introduzidas novamente ao modelo as covariáveis descartadas no passo 2, com o intuito de verificar se elas realmente não contribuem com o ajuste global do modelo.
- **Passo 5:** O passo 5 consiste em retirar novamente as covariáveis uma a uma para verificar se, após todos os passos anteriores, todas elas estão contribuindo para o ajuste global do modelo.

Após a seleção de covariáveis o modelo foi ajustado com as covariáveis que permaneceram no passo 5.

Referências

- ALVARES, D.; LAZARO, H.; GOMEZ-RUBIO, V.; ARMERO, C. Bayesian survival analysis with BUGS. **Statistics in Medicine**, v. 40, n. 12, p. 2975–3020, 30 maio 2021.
- ASSUNÇÃO, J.; GANDOOUR, C.; ROCHA, R. DETER-ing Deforestation in the Amazon: Environmental Monitoring and Law Enforcement. **Applied Economics**, v. 15, n. 2, 2023.
- ASSUNÇÃO, R.; KRAINSKI, E. Neighborhood Dependence in Bayesian Spatial Models. **Biometrical Journal**, v. 51, n. 5, p. 851–869, out. 2009.
- BALAN, T. A.; PUTTER, H. A tutorial on frailty models. **Statistical Methods in Medical Research**, v. 29, n. 11, p. 3424–3454, nov. 2020.
- BANERJEE, S. Frailty modeling for spatially correlated survival data, with application to infant mortality in Minnesota. **Biostatistics**, v. 4, n. 1, p. 123–142, 1 jan. 2003.
- BARONA, E.; RAMANKUTTY, N.; HYMAN, G.; COOMES, O. T. The role of pasture and soybean in deforestation of the Brazilian Amazon. **Environmental Research Letters**, v. 5, n. 2, p. 024002, abr. 2010.
- BESAG, J. Spatial Interaction and the Statistical Analysis of Lattice Systems. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, v. 36, n. 2, p. 192–225, 1 jan. 1974.
- BIVAND, R. S.; GÓMEZ-RUBIO, V.; RUE, H. Spatial Data Analysis with R - **INLA** with Some Extensions. **Journal of Statistical Software**, v. 63, n. 20, 2015.
- CARROLL, K. J. On the use and utility of the Weibull model in the analysis of survival data. **Controlled Clinical Trials**, v. 24, n. 6, p. 682–701, dez. 2003.
- CARROLL, R.; LAWSON, A. B.; FAES, C.; KIRBY, R. S.; AREGAY, M.; WATJOU, K. Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. **Spatial and Spatio-temporal Epidemiology**, v. 14–15, p. 45–54, jul. 2015.
- COLLETT, D. **Modelling Survival Data in Medical Research**. 1. ed. London, UK: Chapman & Hall, 1994.
- COLOSIMO, E. A.; GIOLO, S. R. **Análise de sobrevivência aplicada**. 2. ed. São Paulo, SP: Editora Edgard Blucher, 2024.
- COWLES, M. K. **Applied Bayesian Statistics: With R and OpenBUGS Examples**. New York, NY: Springer New York, 2013. v. 98
- COX, D. R. Regression Models and Life-Tables. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, v. 34, n. 2, p. 187–202, 1 jan. 1972.
- DINIZ, C. G. DETER-B: The New Amazon Near Real-Time Deforestation Detection System. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 8, n. 7, 2015.

- FEARNSIDE, P. M. Deforestation in Brazilian Amazonia: History, Rates, and Consequences. **Conservation Biology**, v. 19, n. 3, p. 680–688, jun. 2005.
- FEARNSIDE, P. M.; LAURANCE, W. F. **Infraestrutura na Amazônia: as lições dos planos plurianuais**. Caderno CRH, v. 25, n. 64, p. 87–98, abr. 2012.
- GELMAN, A.; HWANG, J.; VEHTARI, A. Understanding predictive information criteria for Bayesian models. **Statistics and Computing**, v. 24, n. 6, p. 997–1016, nov. 2014.
- GOVINDARAJULU, U. S.; LIN, H.; LUNETTA, K.; D'AGOSTINO, R. B. Frailty models: Applications to biomedical and genetic studies. **Statistics in Medicine**, v. 30, n. 22, p. 2754–2764, 30 set. 2011.
- HANAGAL, D. D.; DABADE, A. D. Bayesian Estimation of Parameters and Comparison of Shared Gamma Frailty Models. **Communications in Statistics - Simulation and Computation**, v. 42, n. 4, p. 910–931, abr. 2013.
- HOIJTINK, H.; MULDER, J.; VAN LISSA, C.; GU, X. A tutorial on testing hypotheses using the Bayes factor. **Psychological Methods**, v. 24, n. 5, p. 539–556, out. 2019.
- JUNG, S.; POLASKY, S. Partnerships to prevent deforestation in the Amazon. **Journal of Environmental Economics and Management**, v. 92, p. 498–516, nov. 2018.
- JUSYS, T. Fundamental causes and spatial heterogeneity of deforestation in Legal Amazon. **Applied Geography**, v. 75, p. 188–199, out. 2016.
- KAPLAN, E. L.; MEIER, P. Nonparametric Estimation from Incomplete Observations. **Journal of the American Statistical Association**, v. 53, n. 282, p. 457–481, jun. 1958.
- KASS, R. E.; RAFTERY, A. E. **Bayes Factors and model uncertainty**. 29 mar. 1993.
- KINNEBREW, E.; OCHOA-BRITO, J. I.; FRANCH, M.; MILLS-NOVOA, M.; SHOFFNER, E.; SIEGEL, K.; ALDRICH, S. P. Biases and limitations of Global Forest Change and author-generated land cover maps in detecting deforestation in the Amazon. **Plos one**, v. 17, n. 7, p. e0268970, 6 jul. 2022.
- LAVINE, M. L.; HODGES, J. S. On Rigorous Specification of ICAR Models. **The American Statistician**, v. 66, n. 1, p. 42–49, fev. 2012.
- LIKAS, A.; VLASSIS, N.; J. VERBEEK, J. The global k-means clustering algorithm. **Pattern Recognition**, v. 36, n. 2, p. 451–461, fev. 2003.
- LINDGREN, F.; RUE, H. Bayesian Spatial Modelling with *R* - **INLA**. **Journal of Statistical Software**, v. 63, n. 19, 2015.
- MÁQUINA, C. M.; DEMARQUI, F. N.; PRATES, M. O.; SANTOS, N. T. Análise de dados de Sobrevivência espacialmente correlacionados. **Revista Brasileira de Biometria**, v. 37, n. 4, p. 419–434, 20 dez. 2019.
- MARTINO, S.; RUE, H. Case studies in Bayesian computation using INLA. Em: MANTOVAN, P.; SECCHI, P. (Eds.). **Complex Data Modeling and Computationally Intensive Statistical Methods**. Contributions to Statistics. Milano: Springer Milan, 2010. p. 99–114.

MARTINS FILHO, S.; DUARTE, M. L.; VENZON, M. Survival Analysis of the Green Lacewing, *Chrysoperla externa* (Hagen) Exposed to Neem-Based Products. **Agriculture**, v. 13, n. 2, p. 292, 26 jan. 2023.

MARTINS, T. G.; SIMPSON, D.; LINDGREN, FINN.; RUE, H. Bayesian computing with INLA: New features. **Computational Statistics & Data Analysis**, v. 67, p. 68–83, nov. 2013.

MENA, C. F.; BILSBORROW, R. E.; MCCLAIN, M. E. Socioeconomic Drivers of Deforestation in the Northern Ecuadorian Amazon. **Environmental Management**, v. 37, n. 6, p. 802–815, jun. 2006.

MONJARDIN-ARMENTA, S. A.; PLATA-ROCHA, W.; PACHECO-ÂNGULO, C.; FRANCO-OCHOA, C.; RANGEL-PERAZA, J. G. Geospatial Simulation Model of Deforestation and Reforestation Using Multicriteria Evaluation. **Sustainability**, v. 12, n. 24, p. 10387, 11 dez. 2020.

MONTEIRO, A. M. V.; RENNO, C. D.; ALMEIDA, C. A.; VALERIANO, D. M.; SOARES, J. V.; MAURANO, L. E. P.; ESCADA, M. I. S.; AMARAL, S.; PINHEIRO, T. F. **Relatório Técnico-Científico**. 20 maio 2008.

MORAGA, P.; DEAN, C.; INOUE, J.; MORAWIECKI, P.; NOUREEN, S. R.; WANG, F. Bayesian spatial modelling of geostatistical data using INLA and SPDE methods: A case study predicting malaria risk in Mozambique. **Spatial and Spatio-temporal Epidemiology**, v. 39, p. 100440, nov. 2021.

MULLISSA, A.; SAATCHI, S.; DALAGNOL, R.; ERICKSON, T.; PROVOST, N.; OSBORN, F.; ASHARY, A.; MOON, VIOLET.; MELLING, D. LUCA: A Sentinel-1 SAR-Based Global Forest Land Use Change Alert. **Remote Sensing**, v. 16, n. 12, p. 2151, 13 jun. 2024.

NEVES, A. K.; KORTING, T. S.; FONSECA, L. M. G.; ESCADA, M. I. S. Assessment of TerraClass and MapBiomas data on legend and map agreement for the Brazilian Amazon biome. **Acta Amazonica**, v. 50, n. 2, p. 170–182, jun. 2020.

NTZOUFRAS, I. **Bayesian modeling using WinBUGS**. Hoboken, N.J: Wiley, 2009.

PAN, W. **Using Frailties in the Accelerated Failure Time Model**. 2001.

PANWAR, R.; PINKSE, J.; CASHORE, B.; HUSTED, B. W. Why corporate sustainability initiatives fail to reduce deforestation and what to do about it. **Business Strategy and the Environment**, v. 32, n. 8, p. 5121–5127, dez. 2023.

PEREIRA, E. Q. et al. Accelerated failure time frailty model for modeling multiple systems subject to minimal repair. **Applied Stochastic Models in Business and Industry**, v. 40, n. 4, p. 1182–1201, jul. 2024.

RUE, H.; MARTINO, S.; CHOPIN, N. Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, v. 71, n. 2, p. 319–392, 1 abr. 2009.

SALES, V. G.; STROBL, E.; ELLIOTT, R. J. R. Cloud cover and its impact on Brazil's deforestation satellite monitoring program: Evidence from the cerrado biome of the Brazilian Legal Amazon. **Applied Geography**, v. 140, p. 102651, mar. 2022.

SANTOS, A. M. D.; SILVA, C. F.; ALMEIDA JUNIOR, P.; RUDKE, A. P.; MELO, S. N. Deforestation drivers in the Brazilian Amazon: assessing new spatial predictors. **Journal of Environmental Management**, v. 294, p. 113020, set. 2021.

SCHMALZ, X.; BIURRUN MANRESA, J.; ZHANG, L. What is a Bayes factor? **Psychological Methods**, v. 28, n. 3, p. 705–718, jun. 2023.

SILVA-JUNIOR, C. H. L.; SILVA, F. B.; ARISI, B. M.; MATAVELI, G.; PESSOA, A. C. M.; CARVALHO, N. S.; REIS, J. B. C.; SILVA JUNIOR, A. S.; MOTTA, N. A. C. S.; SILVA, P. V. M. E.; RIBEIRO, F. D.; SIQUEIRA-GAY, J.; ALENCAR, A.; SAATCHI, S.; ARAGAO, L. E. O. C.; ANDERSON, L. O.; MELO, M. Brazilian Amazon indigenous territories under deforestation pressure. **Scientific Reports**, v. 13, n. 1, p. 5851, 10 abr. 2023.

SONTER, L. J.; HERRERA, D.; BARRET, D. J.; GALFORD, G. L.; MORAN, C. J.; SOARES-FILHO, B. S. Mining drives extensive deforestation in the Brazilian Amazon. **Nature Communications**, v. 8, n. 1, p. 1013, 18 out. 2017.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE, A. Bayesian Measures of Model Complexity and Fit. **Journal of the Royal Statistical Society Series B: Statistical Methodology**, v. 64, n. 4, p. 583–639, 1 out. 2002.

THERNEAU, T. M. **Survival: Survival Analysis**. 22 jun. 2001. Disponível em: <<https://CRAN.R-project.org/package=survival>>. Acesso em: 3 fev. 2025

UMETA, A. K.; YERMOSA, S. F.; DUFERA, A. G. Bayesian parametric modeling of time to tuberculosis co-infection of HIV/AIDS patients at Jimma Medical Center, Ethiopia. **Scientific Reports**, v. 12, n. 1, p. 16475, 1 out. 2022.

VHUDZIJENA, M. **Modelling mortality heterogeneity using health trajectories and multimorbidity**. [s.l.] UNSW Sydney, 2023.

WANG, X.; RYAN, Y. Y.; FARAWAY, J. J. **Bayesian Regression Modeling with INLA**. jan. 2018.

CAPÍTULO 2: MODELAGEM ESPACIAL DO DESMATAMENTO NO SUL DO AMAZONAS: UMA ABORDAGEM UTILIZANDO ANÁLISE DE SOBREVIVÊNCIA

O capítulo a seguir foi redigido nas normas da revista *Acta Amazônica*. A seção a seguir aborda uma contextualização do problema associado ao desmatamento e uma descrição de todos os passos para coleta de dados, análise e interpretação dos resultados obtidos, com indicações de como eles se mostram relevantes para o estudo das dinâmicas associadas ao desmatamento na região.

2.1.1. Resumo

O desmatamento é um problema mundial que compromete ciclos ecológicos em grandes ecossistemas, e que vem sendo estudado e monitorado constantemente. Desse modo, o presente trabalho objetivou aplicar técnicas para a análise espacial do desmatamento e análise de sobrevivência, com o intuito de verificar o efeito de covariáveis de interesse e inserir a dependência dos tempos observados na modelagem dos dados de desmatamento. Os dados anuais de desmatamento utilizados foram obtidos por meio do *Mapbiomas* e processados junto às demais covariáveis. A análise de sobrevivência foi aplicada por meio da estimação não paramétrica, além de modelos paramétricos e do modelo de fragilidade. Todas as covariáveis se mostraram significativas na modelagem do desmatamento, porém as covariáveis pastagem, unidades de conservação, terras indígenas e distância do pixel até rodovias, demonstraram uma importância maior para essa finalidade na modelagem do desmatamento. Além disso, foi possível verificar que o desmatamento é diretamente afetado pela distância da área às estradas, e, em unidades de conservação, o tempo esperado até região ser desmatada é menor, assim como em terras indígenas. Os valores associados a efeitos aleatórios de

cada grupo demonstraram o quão frágil é a região à ocorrência do desmatamento, independentemente do valor observado para as covariáveis, por meio do modelo de fragilidade.

Palavras-chave: Modelos de fragilidade compartilhada; Modelagem Bayesiana; Modelo ICAR; Análise de vizinhança; Floresta Amazônica.

2.1.2. Abstract

Deforestation is a global issue that, due to its various environmental impacts on large ecosystems, has been continuously studied and monitored. Thus, this study aimed to apply spatial analysis techniques to deforestation and survival analysis to assess the effect of relevant covariates and how random effects behave in this region. The annual deforestation data used were obtained from MapBiomas and processed alongside other variables. Survival analysis was conducted using non-parametric estimation, parametric models, and the frailty model. All variables were found to be significant in modeling deforestation; however, pasture, conservation units, indigenous lands, and the distance from a pixel to roads proved to be the most important factors in this context. Furthermore, it was observed that deforestation is directly influenced by the distance of an area from roads, and in conservation units, the expected time until the region is deforested is shorter, as is the case in indigenous lands. The random effect values for each group demonstrated how vulnerable the region is to deforestation, regardless of the observed covariate values, as indicated by the frailty model.

Keywords: Shared frailty models; Bayesian modeling; ICAR model; Neighborhood analysis; Amazon Rainforest.

2.2. Introdução

A Amazônia é um bioma que exerce um papel crucial no Brasil e no mundo devido a vários fatores ambientais, climáticos e ecológicos. A Floresta Amazônica abrange cerca de 10% da biodiversidade da terra segundo dados da plataforma Global Forest Watch (GFW), além de sequestrar uma grande quantidade de dióxido de carbono (CO_2) processo natural de armazenamento de carbono por plantas, algas entre outros, o que ajuda a estabilizar o clima do planeta.

Os “rios voadores”, ou flutuantes, são caracterizados pelo grande fluxo de vapor de água que circula da região amazônica para o Centro-Oeste e Sudeste e outras regiões, devido a diferença de temperatura característico do clima equatorial (Fearnside and Laurance 2012). Sob essa perspectiva, a Floresta Amazônica ajuda a controlar a frequência de chuvas na região, além de possuir efeito direto no ciclo do carbono (Flores et al. 2024; Zemp et al. 2017). Fearnside (2005) também afirma que a ausência desse fenômeno na região comprometeria seriamente o ciclo de chuvas em outras regiões no Brasil e em países vizinhos, e as cidades se assemelhariam a um deserto.

No entanto, o bioma vem sofrendo pressões cada vez maiores com o aumento do desmatamento na região. O desmatamento é um dos problemas mundiais e que foi alvo de várias iniciativas para o controle e a boa gestão ambiental (Moeliono et al. 2020, Neves et al. 2020, Assunção et al. 2023). É evidente, portanto, a importância de se buscar descrever como se comportam as dinâmicas associadas ao desmatamento e como relaciona-se ao impacto ambiental gerado por tais problemas.

Os modelos estatísticos desempenham um papel importante no estudo do desmatamento, uma vez que as ocorrências podem estar associadas a uma natureza complexa. Além disso, técnicas robustas de modelagem podem ajudar a identificar áreas de preservação e áreas que possuem maior probabilidade de serem desmatadas com relação a causas diretas e indiretas ao desmatamento (Ball et al. 2022; Miranda et al. 2018; Monteiro et al. 2008).

Diversos métodos de análise foram empregados com o objetivo de se estudar o desmatamento e suas dinâmicas, e melhor descrever o comportamento associado ao manejo da terra para controle e preservação. Alguns métodos estatísticos como regressão logística (Saha et al. 2020), regressão geograficamente ponderada (GWR) (Jusys 2016), regressão de Poisson (Achcar et al. 2011), e métodos de machine learning (Dias et al., 2024; Dominguez et al., 2022), foram propostos para modelar o desmatamento. Bell and Irwin (2002) e Greenberg et al. (2005) apresentaram alguns dos primeiros estudos com o propósito de modelar o desmatamento por meio da análise de sobrevivência. Xin et al. (2023) utilizaram análise de sobrevivência na modelagem do desmatamento com enfoque na utilização do modelo de riscos proporcionais de Cox. Essa mesma abordagem foi utilizada por Sales et al. (2022) para avaliação de como a cobertura por nuvens afeta o monitoramento do desmatamento por satélites.

A análise de sobrevivência é um ramo de estudo que busca analisar o tempo até a ocorrência de um evento de interesse. Sua principal característica é a presença de censuras, representadas por indivíduos que não experimentaram o evento (Colosimo and Giolo 2024). Portanto, a possibilidade de inserir na análise os indivíduos censurados se configura uma vantagem da utilização da análise de sobrevivência em relação aos demais métodos.

A partir da análise espacial do desmatamento, é possível incorporar ao modelo efeitos aleatórios de covariáveis não mensuradas que podem ser de grande relevância para a análise de locais de interesse, uma vez que sua interpretação não é condicionada às covariáveis, mas sim aos demais grupos analisados. Esses grupos representam estruturas de áreas bem definidas espacialmente, em que a dependência espacial faz com que esses grupos sejam correlacionados ao tempo até o desmatamento da unidade de área desmatada. Com esse propósito, a análise do desmatamento pode ser conduzida por meio dos modelos de fragilidade, uma importante classe dos modelos de análise de sobrevivência.

Essa abordagem se diferencia por permitir a análise de efeitos que afetam diretamente ou indiretamente o desmatamento, e por possibilitar a análise espacial do território por meio de efeitos aleatórios. Esses efeitos aleatórios podem ser entendidos como efeitos associados à covariáveis latentes aos dados.

Essa análise permite uma identificação de áreas frágeis (regiões em que o termo de fragilidade é maior, e, conseqüentemente, possui maior propensão ao desmatamento devido a fatores aleatórios) ao evento de interesse. Essa aplicação é de grande relevância para um cenário de preservação, estudo e monitoramento do desmatamento. Com isso, essa análise traz não somente informações descritivas, mas informações que não são detectadas diretamente pela observação das ocorrências de desmatamento.

A partir desse contexto, de maneira geral, este estudo possui o objetivo de avaliar os efeitos das covariáveis socioeconômicas e ambientais na ocorrência do desmatamento utilizando a análise de sobrevivência e análise espacial da região de estudo por meio do modelo de fragilidade no contexto multivariado. Especificamente, os objetivos são a utilização de modelos de fragilidade em conjuntura a uma análise de sobrevivência para

modelar o tempo até a ocorrência do desmatamento, bem como a identificação de covariáveis que influenciam significativamente o risco de desmatamento.

2.3. Material e métodos

2.3.1. Área de estudo

A área de estudo compreende a porção sul do estado do Amazonas, pertencente à uma região conhecida como “arco do desmatamento” (Butt et al., 2023) e ao tipo floresta ombrófila densa, representada na Figura 1. O período considerado para o estudo foi de 1987 a 2023. O território divide uma região de áreas desmatadas com uma porção considerável da Floresta Amazônica e com grandes chances para a ocorrência do desmatamento nos próximos anos.

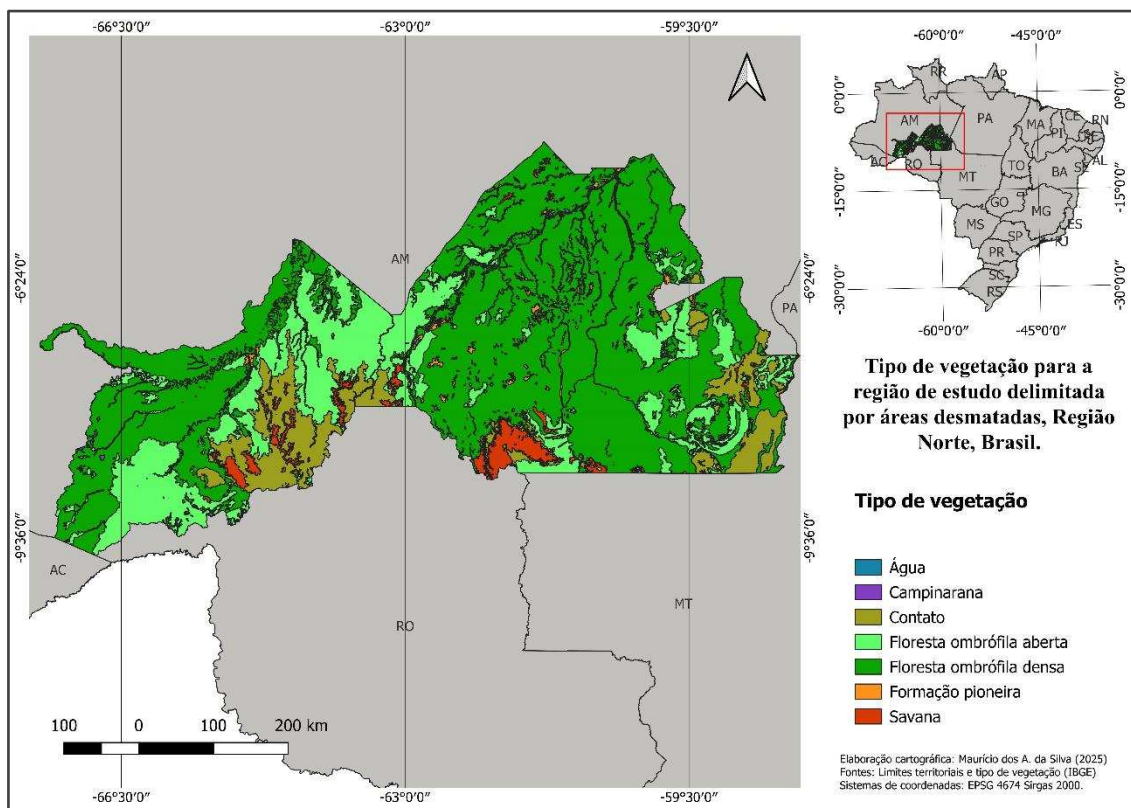


Figura 1 - Mapa da região de estudo delimitado na região norte do Brasil, por tipo de vegetação, com as coordenadas geográficas correspondentes.

2.3.2. Dados de desmatamento

Os dados foram obtidos a partir do MapBiomas Collection 9.0 e IBGE, e estão disponíveis para livre acesso por meio do Google Earth Engine. Esses dados foram processados no QGIS Desktop e no software R (R Core Team 2024). Para cada ano, uma imagem foi processada com áreas de pixels desmatados e não desmatados. A variável resposta foi composta pela junção de todas essas imagens contendo variáveis binárias representadas pela presença de desmatamento na região do pixel em cada ano específico, sendo que 1 pixel corresponde a 1 km² da região (Sales et al. 2022; Souza et al. 2020; Santos et al. 2021).

A escala de reclassificação de cada pixel foi definida com base em observações feitas acerca das categorias dos pixels diretamente no site do Google Earth Engine, e a atribuição da nova escala reclassificada a todos os pixels foi conduzida por meio do Qgis Desktop 3.36.1. A Tabela 3 representa as regras de reclassificação utilizadas para o desmatamento.

Tabela 3 - Regras de reclassificação dos pixels para o desmatamento anual, com identificação da classe referente às áreas desmatadas e não desmatadas.

Escala do pixel	Categoria Reclassificada	Descrição
103 a 116	1	Presença de desmatamento
116 a 733	0	Área não desmatada

Fonte: O autor (2025).

Nesse sentido, após a reclassificação dos pixels contendo os dados de desmatamento, foi conduzido o processo de poligonalização de todas as camadas anuais. As imagens contendo os dados de desmatamento foram agregados de modo a conter apenas o tempo até o primeiro desmatamento naquela região, em que também é necessário para apurar as regiões censuradas.

2.3.3. Covariáveis

As covariáveis consideradas para a modelagem do desmatamento foram unidades de conservação, terras indígenas, pastagem, potencialidade agrícola, áreas urbanas, distância às rodovias, distância às hidrovias, distância à mineração e população. Todas as covariáveis foram extraídas a partir do site do MapBiomas e IBGE, a partir de camadas em formatos raster e polígonos para a camada de polígonos principal contendo o desmatamento, e esse procedimento foi conduzido a partir do Qgis Desktop 3.36.1, com o intuito de unir os dados de desmatamento às covariáveis presentes nas demais bases de dados. As covariáveis passaram então pelos procedimentos de extração, recorte, reclassificação, poligonalização, e junção com os dados de desmatamento.

Seja a região de um pixel da imagem correspondente a 1 km^2 . Sob essa região, é possível que ocorra a presença de duas ou mais classes da mesma covariável. Um exemplo pode ser observado para a variável pastagem, em que a região de um pixel pode estar sujeita a 3 tipos de pastagem. O critério adotado para esse caso foi considerar a classe que possui maior cobertura de área e atribuí-la ao pixel associado. Além disso, foram consideradas áreas protegidas as regiões dentro de unidades de conservação.

2.3.4. Métodos não paramétricos

O estimador de Kaplan-Meier foi aplicado para a estimação da curva de não desmatamento e comparação de cada nível associado à cada covariável (Colosimo and Giolo 2024). A curva de não desmatamento representa as probabilidades associadas aos tempos analisados, em que se avalia a probabilidade de que a área não seja desmatada. Nesse cenário, foi possível incorporar a informação de falhas e censuras nas análises e comparar os níveis de cada covariável com relação ao tempo até a ocorrência do evento

de interesse (Kaplan and Meier 1958). O objetivo foi comparar o número de indivíduos que falharam com o total de indivíduos sob risco.

Seja q_j a probabilidade de falha no intervalo $(t_{j-1}, t_j]$ para $j = 1, 2, \dots, k$. A fórmula geral para o estimador de Kaplan-Meier é dada por:

$$\hat{S}(t) = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j} \right)$$

em que representa o total de indivíduos sob risco, ou seja, os que não falharam e não foram censurados até t_{j-1} e d_j representa o total de falhas em $(t_{j-1}, t_j]$. Com isso, a probabilidade de sobrevivência foi estimada considerando também os indivíduos censurados.

As covariáveis quantitativas foram categorizadas pelo valor mediano. Essa categorização significa que, pixels com valores abaixo e acima do valor mediano calculado para a covariável, foram alocados em classes diferentes e as curvas de sobrevivência foram estimadas de maneira análoga às demais covariáveis pelo estimador de Kaplan-Meier. Além disso, o teste Log-rank foi empregado para comparação entre as curvas estimadas, com o intuito de identificar as covariáveis que apresentam diferenças estatisticamente significativas na curva de sobrevivência de cada nível observado.

2.3.5. Métodos paramétricos

Os modelos paramétricos foram aplicados aos dados de desmatamento para a estimação das curvas de sobrevivência, com o intuito de inserir a informação presente nas covariáveis de interesse na modelagem, uma vez que esse procedimento não é viabilizado

em técnicas não paramétricas (Colosimo and Giolo, 2024). Nesse sentido, todos os procedimentos referentes ao processamento dos dados foram mantidos, com exceção da categorização das covariáveis quantitativas.

Os modelos considerados para o estudo do desmatamento foram: Exponencial, Weibull, Log-normal e Log-logístico. Tais distribuições representam os principais modelos paramétricos utilizados em estudos análise de sobrevivência com esse mesmo propósito (Colosimo and Giolo, 2024).

O procedimento de seleção de covariáveis foi conduzido com a aplicação de uma adaptação do método de Collett (Collett 1994) considerando o DIC e o fator de Bayes para tomada de decisão quanto ao melhor modelo. As análises foram conduzidas por meio do software R (R Core Team 2025) com a utilização do pacote *survival* (Therneau, 2001) e com o auxílio do pacote *INLA* (Held et al. 2010; Rue et al. 2009).

2.3.6. Os modelos de fragilidade

O modelo de fragilidade foi aplicado para a modelagem do desmatamento, com o intuito de inserir no modelo a heterogeneidade não observada (Colosimo and Giolo, 2024; Vaupel et al. 1979; Balan and Putter 2020). Os modelos de fragilidade podem também ser aplicados tanto no contexto de modelos de riscos proporcionais, quanto no contexto de modelos de tempo de falha acelerados (AFT). Para o modelo Weibull (AFT) (Rue et al. 2009), a fórmula geral do modelo de fragilidade é dada por:

$$f(t) = \alpha t^{\alpha-1} \lambda \exp(-\lambda t^\alpha)$$

em que $\alpha > 0$ representa o parâmetro de forma e $\lambda > 0$ representa o parâmetro de escala. Desse modo, $\lambda = \exp(\eta)$ e $\eta = \mathbf{x}_{it}'\boldsymbol{\beta} + w_{it}$, em que η representa o preditor linear e w_{it} representa o termo que caracteriza o modelo de fragilidade. Várias distribuições podem

ser assumidas para a modelagem do termo de fragilidade, dentre elas a distribuição gama, comum para esse propósito. Neste estudo será utilizada a distribuição normal para essa finalidade.

Nesse sentido, a função de verossimilhança do modelo pode ser expressa por:

$$L(\mu, \sigma) = \prod_{l=1}^L \prod_{i=1}^{n_l} [f(t_{l,i} | x_{li})]^{\delta_{li}} S(t_{l,i} | x_{li})^{1-\delta_{li}}$$

em que δ_{li} representa o indicador de falha ou censura, em um cenário formado por L grupos e n_l indivíduos em cada grupo (Hanagal and Dabade, 2013; Umeta et al. 2022; Banerjee 2003).

As análises conduzidas para o ajuste do modelo de fragilidade foram processadas com o auxílio do pacote *INLA* (Held et al. 2010; Rue et al. 2009), por meio do software R (R Core Team 2025).

Nos modelos de tempo de falha acelerados (AFT), o fator de aceleração pode ser definido como o valor do coeficiente estimado pelo modelo já transformado na escala de interpretação. Nesse momento, esses valores representam visivelmente o percentual de aceleração ou desaceleração no tempo até a região do pixel ser desmatada.

A escolha dos grupos a serem inseridos para compor os efeitos aleatórios, deve estar associada ao fator pelo qual os tempos são dependentes. Nesse sentido, o método K-Means foi utilizado para a identificação de grupos na dispersão espacial do desmatamento para utilização no modelo de fragilidade (Govindarajulu and Bedi 2022).

As análises foram conduzidas por meio da abordagem bayesiana, com a utilização de distribuições *a priori* vagas. As seguintes distribuições foram definidas para representação da informação *a priori*:

$$\begin{aligned}\beta &\sim N(0; 10^3) \\ \alpha &\sim \log - \text{gamma}(1, 10^8) \\ \phi &\sim \log - \text{gamma}(1, 10^8) \\ (w_l | \mathbf{w}_{(-l)}) &\sim N\left(\bar{w}_l; \frac{1}{(\phi n_l)}\right), \quad l = 1, \dots, L\end{aligned}$$

em que os valores β são representados por uma distribuição normal, com média 0 e variância 1000. Ambos os parâmetros relacionados à da distribuição log-gamma foram descritos com parâmetro de forma igual a 1 e parâmetro de escala igual a 10^8 . Para os termos definidos para w_l e associados ao modelo ICAR, \bar{w}_l representa a média condicional de w_l e $\frac{1}{(\phi n_l)}$ a sua variância condicional, sendo ϕ o parâmetro de precisão associado e n_l o número de vizinhos identificados na matriz de vizinhanças.

O modelo ICAR (*Intrinsic Conditional Autoregressive Model*) foi aplicado para descrever o termo de efeito aleatório (Besag 1974, Assunção and Krainski 2009, Lavine and Hodges 2012), com relação ao agrupamento pré-determinado. Além disso, para a criação da estrutura de vizinhanças definida por meio da inserção da matriz de adjacências no modelo, foi considerado o *queens case*, em que um polígono deve ser considerado vizinho mesmo se compartilhar apenas um vértice representado por um único ponto em sua fronteira de vizinhança (Assunção and Krainski 2009, Azevedo et al. 2020).

2.3.7. Adequação do modelo

Os resíduos de Cox-Snell foram utilizados para a verificação da qualidade do ajuste do modelo (Wang et al. 2018). Embora sejam populares no contexto do ajuste paramétrico da análise de sobrevivência clássica, uma análise dos resíduos de Cox-Snell pode ser conduzida no contexto bayesiano para verificação da aproximação da distribuição a

posteriori com relação à família escolhida para compor a verossimilhança (Bivand and Gómez-Rubio, 2021; Gashu and Aguade, 2024). Para a adequação do modelo, os resíduos devem se aproximar da função de risco acumulado da distribuição exponencial padrão (Ashine et al. 2021).

Foi utilizado também o índice de Kullback-Leibler (KLD), (Bivand et al. 2015, Silva et al. 2020, Vhudzijena 2023), para a avaliação da aproximação da distribuição *a posteriori* encontrada com a verdadeira distribuição *a posteriori*. Para que se tenha uma boa aproximação, o índice de Kullback-Leibler (KLD) deve estar próximo a zero.

Para comparação de diferentes modelos, foram utilizados o DIC (Deviance information criterion) e o fator de Bayes (Hoijtink et al. 2019, Schmalz et al. 2023). Essas métricas foram utilizadas nos procedimentos de seleção de covariáveis e na comparação de modelos com relação ao termo de fragilidade.

2.4. Resultados

Após a seleção de covariáveis, utilizando uma adaptação do método de Collett, a única covariável áreas urbanas.

2.4.1. Estimador de Kaplan-Meier

Pode ser observado na Figura 2 que áreas em unidades de conservação possuem menores probabilidades de desmatamento, o que indica que o tempo esperado até a área do pixel ser desmatada é maior em comparação à áreas não protegidas em todo o período de estudo, o que é esperado, pois as unidades de conservação são protegidas (Figura 2a). De maneira análoga, é possível concluir que todas as covariáveis apresentaram diferenças

significativas nas probabilidades de desmatamento para pelo menos um nível com relação aos demais, pelo teste Log-rank (Figura 2b, c, d).

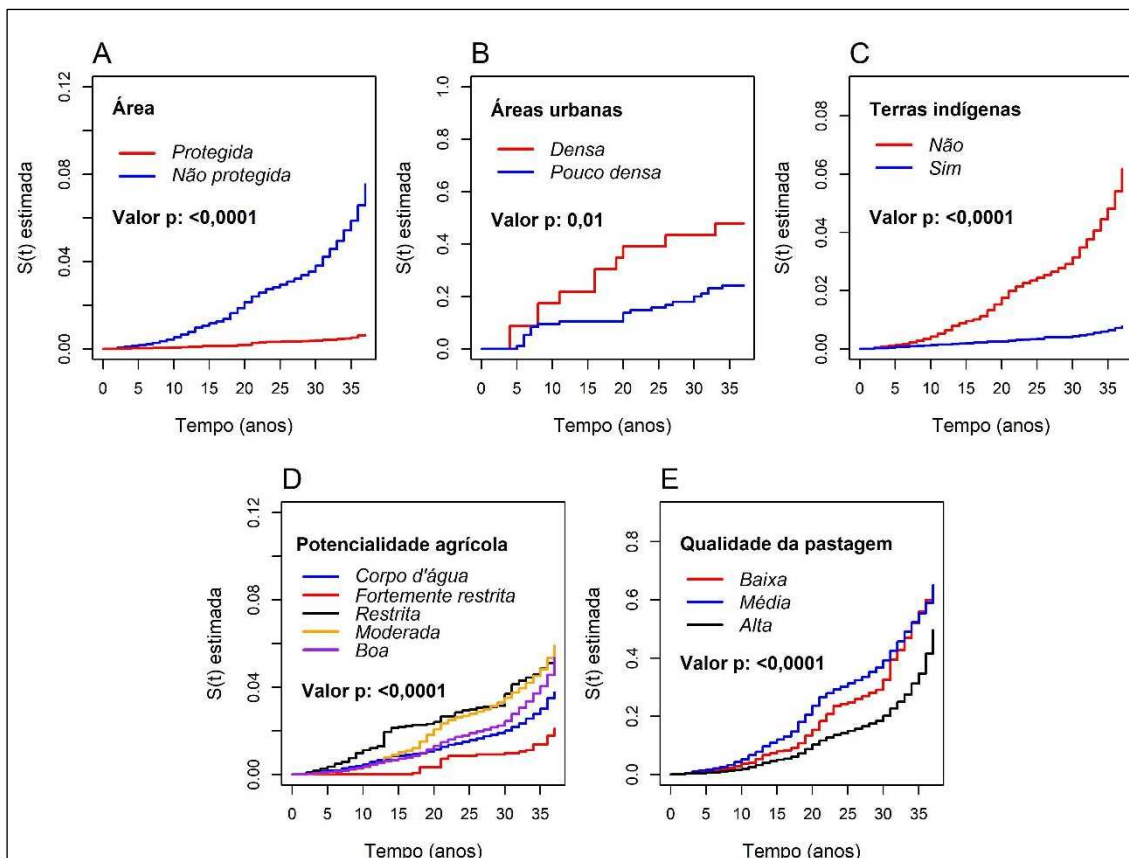


Figura 2 - Curvas para as probabilidades de desmatamento estimadas pelo método não paramétrico de Kaplan-Meier para: (A) Unidades de conservação, (B) Áreas urbanas, (C) Terras indígenas, (D) Potencialidade agrícola e (E) Pastagem; p-valor referente ao teste Log-rank para comparação das curvas estimadas.

As covariáveis quantitativas distâncias da área desmatada até rodovias, hidrovias e pontos de mineração em conjunto à variável população do município, podem ser analisadas de igual forma quando categorizadas. As curvas para as probabilidades de desmatamento para as covariáveis quantitativas categorizadas pelo seu valor mediano podem ser visualizadas na Figura 3.

Na Figura 3 pode ser observado que todas as covariáveis quantitativas categorizadas, pelo seu valor mediano, apresentaram diferenças significativas nas

probabilidades de desmatamento. Esse resultado também pode ser avaliado pelo resultado do teste Log-rank para cada covariável.

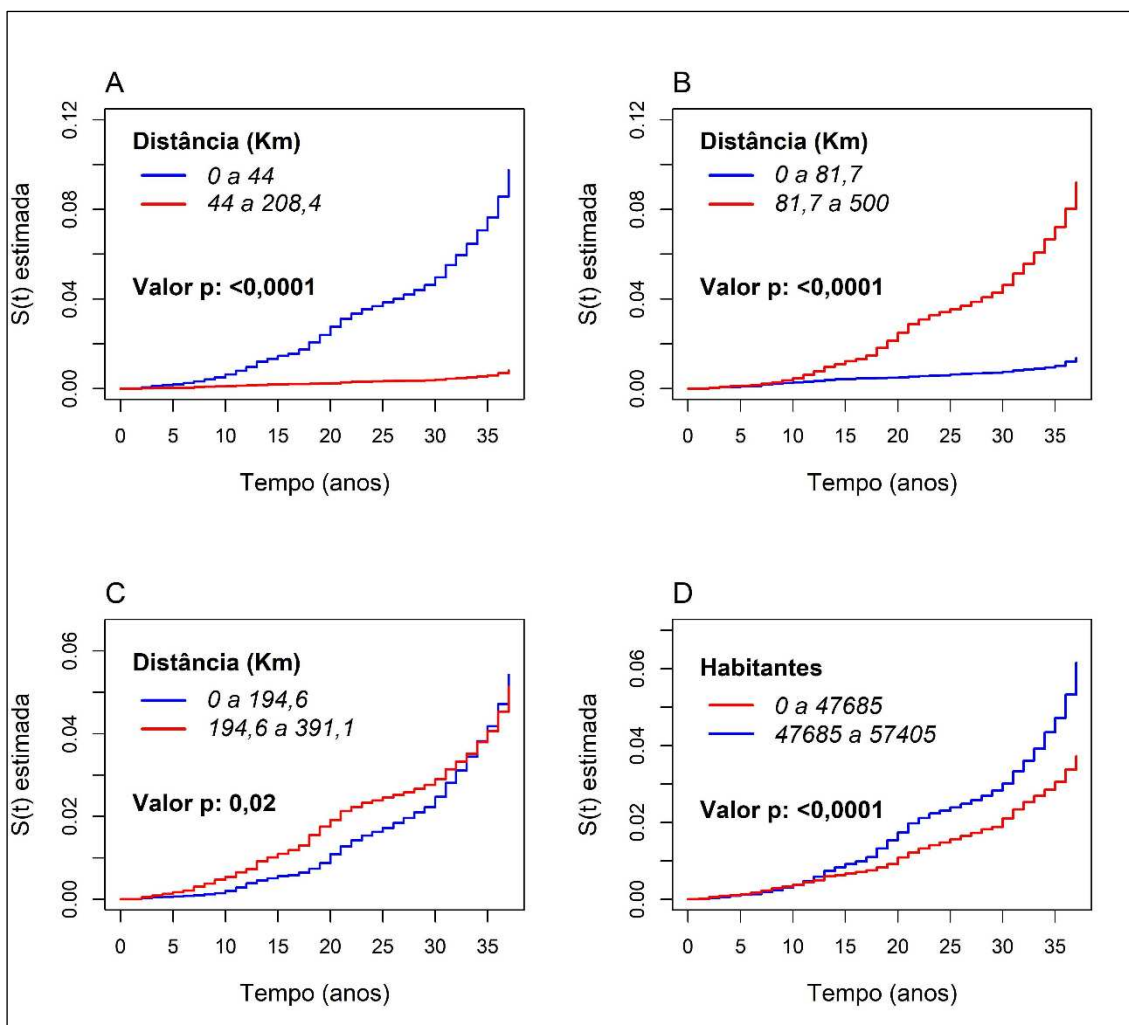


Figura 3 - Curvas para as probabilidades de desmatamento estimadas pelo método não paramétrico de Kaplan-Meier para: (A) Distância até rodovias, (B) Distância até hidrovias, (C) Distância à mineração e (D) População, p-valor referente ao teste Log-rank para comparação das curvas de sobrevivência.

2.4.2. Análise da distribuição *a posteriori* a partir dos resíduos de Cox-Snell

De acordo com a avaliação dos resíduos de Cox-Snell dispostos na

Figura 3b, o modelo Weibull é o mais adequado, pois os resíduos de Cox-Snell estimados a partir dessa família, se aproximaram mais da função de risco acumulado da

distribuição exponencial padrão. Por esse motivo, o modelo Weibull é o que melhor representa a relação entre o desmatamento e as covariáveis, fazendo com que a estimação dos efeitos associados as covariáveis por meio desse modelo seja apropriada. Isso indica que, a escolha dessa família de distribuições para a representação da verossimilhança é mais adequada e representa melhor os dados de desmatamento em conjunto às covariáveis em questão. A consequência prática desse resultado é que a escolha do modelo Weibull para o ajuste do modelo permite a estimação e conclusões robustas quanto à associação das covariáveis ao desmatamento na região.

Para os demais modelos avaliados, os resíduos de Cox-Snell se distanciaram da função de risco acumulado da distribuição exponencial padrão (Figura 3a, c, d).

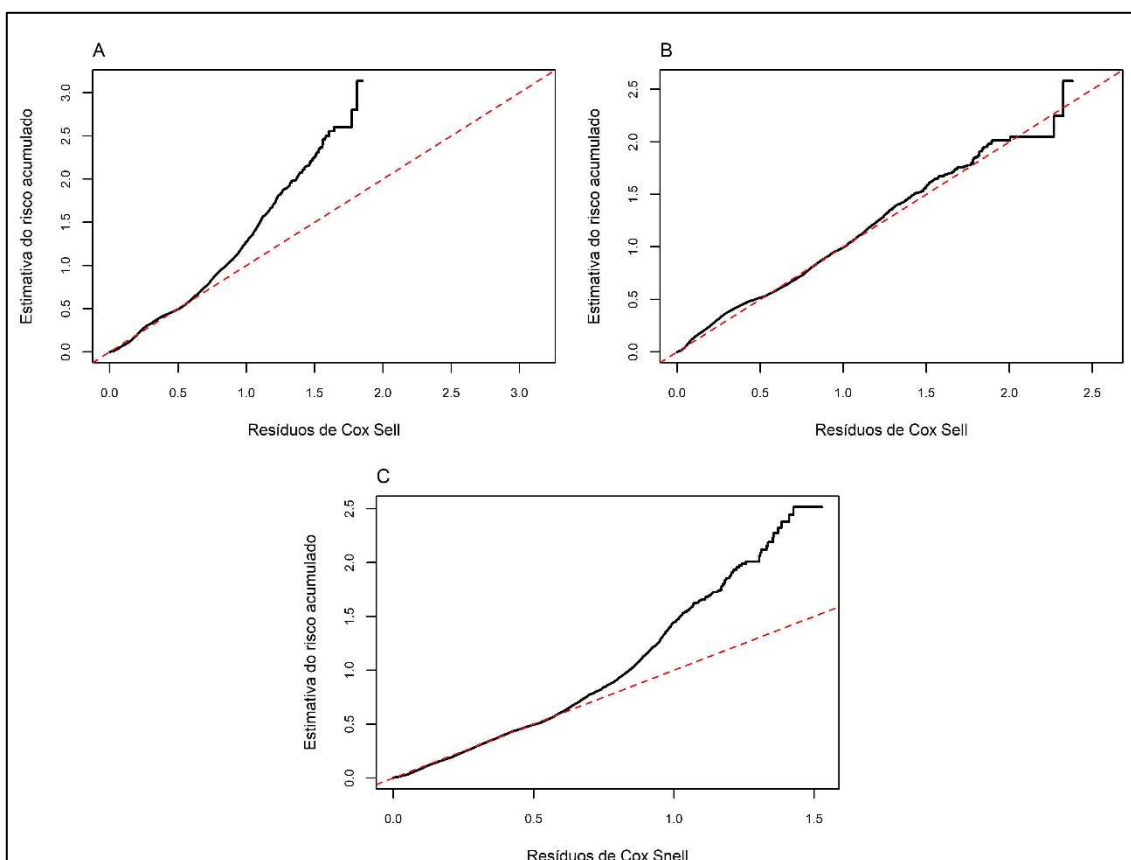


Figura 3 - Resíduos de Cox-Snell para avaliação da adequação geral da distribuição α posteriori aos dados, representada pelos modelos: A) Exponencial, B) Weibull e C) Log-normal.

2.4.3. Resultados para o modelo de fragilidade

Na Tabela 4 são comparados os modelos com e sem o termo de fragilidade. O modelo que inclui o efeito aleatório apresentou melhor desempenho, evidenciado por um menor valor no DIC e um expressivo valor para o fator de Bayes, demonstrando sua superioridade em relação ao modelo sem fragilidade.

Tabela 4 - Tabela para a comparação dos modelos com o termo de fragilidade e sem o termo de fragilidade.

Modelo	DIC	Log-Verossimilhança	Fator de Bayes
Com o efeito aleatório	12166	-6183,45	>100
Sem o efeito aleatório	13564	-6865,23	-

Fonte: O autor (2025).

Isso indica que a estrutura espacial e a informação de vizinhança dos grupos acrescentaram informação ao modelo, uma vez que o ajuste anterior é melhorado com a adição do termo de fragilidade. Esse impacto é causado pela dependência espacial dos dados de desmatamento.

Os resultados para o ajuste com a adição do termo de fragilidade estão dispostos na Tabela 5. Um fator de aceleração de valor 1, não causa nenhum impacto no tempo até o desmatamento pelo seu caráter multiplicativo, o que demonstra que o coeficiente associado à covariável é não significativo.

É possível ver que a maioria das covariáveis se mostraram significativas, com um intervalo de credibilidade que não contém o valor 1, com 95% de probabilidade, o que indica a importância dessas covariáveis para o modelo. O índice de Kullback-Leibler (KLD), apresentou valor muito próximo a zero para todos os coeficientes associados às

covariáveis do modelo, o que indica que o método apresentou uma boa aproximação para esses parâmetros.

Tabela 5 - Coeficientes do modelo de fragilidade ajustado utilizando a distribuição Weibull para a verossimilhança, juntamente com o intervalo de credibilidade com os limites inferiores e superiores a 5% de probabilidade, o valor médio e o índice KLD (Kullback-Leibler divergence).

Variável	$-\zeta \cdot \sigma$	FA	sd	IC _{2,5%}	IC _{97,5%}	KLD
Intercepto	0,4857	1,6176	0,0390	1,5715	1,6653	0
Unid. de conservação ²	0,6374	1,8798	0,0575	1,8018	1,9612	0
Terras indígenas ³	0,5417	1,7098	0,0724	1,6210	1,8035	0
Pastagem	-0,3598	0,7003	0,0109	0,6947	0,7059	0
PA - Boa	0,0190	1,0190	0,0284	0,9978	1,0405	0
PA - Moderada	-0,0741	0,9292	0,0397	0,9024	0,9568	0
PA- Corpo d'água	-0,3942	0,6768	0,0965	0,6303	0,7267	0
PA - Fortemente restrita	0,1657	1,1783	0,1794	1,0324	1,3449	0
Distância à rodovias ¹	0,2816	1,3216	0,0228	1,2996	1,3441	0
Distância à hidrovias ¹	0,0061	1,0060	0,0268	0,9864	1,0261	0
Distância à mineração ¹	0,1165	1,1222	0,0239	1,1027	1,1422	0
População municipal ¹	-0,0951	0,9102	0,0178	0,8983	0,9222	0

¹Variáveis padronizadas; ζ = Coeficientes do modelo de fragilidade; σ = scale, PA = Potencialidade Agrícola; ²Áreas dentro de unidades de conservação; ³Áreas dentro de terras indígenas; sd = desvio padrão; FA = Fator de aceleração.

Fonte: O autor (2025).

Para a variável unidades de conservação, o fator de aceleração ($\exp[-\zeta \cdot \sigma]$) associado foi 1,8798. Dessa maneira, é possível ver que o fato de a região ser uma unidade de conservação aumenta o tempo esperado até o desmatamento do pixel em 87,98% com relação á áreas que não estão presentes em unidades de conservação.

O fato de o pixel estar presente em uma terra indígena, faz com que o tempo esperado até o desmatamento seja aumentado em 70,98%, com um fator de aceleração de 1,7098. Diversos fatores também podem ser elencados quanto a esse efeito, mas é também um resultado esperado devido a maiores dificuldades de se explorar essas terras.

Com relação à covariável pastagem, é possível verificar que o aumento em um nível nessa covariável faz com que o tempo esperado até o desmatamento seja reduzido em 29,97%, mantendo-se fixas todas as demais covariáveis (Tabela 5).

O nível que indica a potencialidade agrícola fortemente restrita na região, está associado a um aumento no tempo esperado até o desmatamento de cerca de 17,83% com relação à região com potencialidade agrícola restrita (Tabela 5). O coeficiente da classe referente a corpo d'água continental, pode ser explicado pelo fato de que a maior cobertura nessas regiões é considerada corpo d'água, o que não impede o desmatamento no restante do pixel ao redor dessas áreas. A presença da maior sobreposição caracterizada como corpo d'água, diminui o tempo até o desmatamento em 32,32%.

A variável distância até estradas, indica a distância euclidiana do pixel até a rodovia mais próxima. A padronização pela média e desvio padrão, faz com que a interpretação seja alterada para essa covariável.

O desvio padrão para a covariável distância até rodovias foi 42997,31 metros, o que representa aproximadamente 43 quilômetros, que é correspondente a alterar em uma unidade a variável padronizada. Portanto, ao se distanciar em 43 quilômetros de rodovias, o tempo esperado até a região ser desmatada é aumentado em 32,16% daquele observado em regiões próximas a rodovias, mantidas constantes todas as demais covariáveis (Tabela 5).

A variável distância até hidrovias embora tenha se mostrado significativa no modelo sem presença dos efeitos aleatórios para cada grupo, para essa abordagem o efeito dessa covariável não se mostrou significativo, com um intervalo de credibilidade contendo o valor 1.

A variável população do município também se mostrou significativa para o modelo, mesmo na presença dos efeitos aleatórios. O desvio padrão para essa covariável observada foi de 15387 habitantes. Nesse sentido, a cada aumento de 15387 habitantes no município, espera-se uma diminuição de 8,98% no tempo esperado até a região ser desmatada, mantidas constantes todas as demais covariáveis (Tabela 5).

Finalmente, a variável distância até o ponto mais próximo de mineração, também calculada pela distância euclidiana do pixel até o ponto mais próximo de acordo com dados disponíveis no MapBiomas, apresentou um desvio padrão de 82117,82 metros. Isso indica que, a cada 82117,82 metros de pontos de mineração, espera-se que o tempo esperado até o desmatamento do pixel seja acrescido de 12,22% com relação às regiões próximas a pontos de mineração, mantendo-se, novamente, fixas as demais covariáveis.

As distribuições marginais, Figura 4, indicam visualmente o nível de importância da covariável para explicar o desmatamento por meio de suas medidas de posição e dispersão. Desse modo, distribuições com maiores médias em módulo e com menor variabilidade indicam as covariáveis mais importantes com relação às demais para explicar o desmatamento. As distribuições marginais para cada um dos parâmetros associados às covariáveis, estão representados na Figura 4:

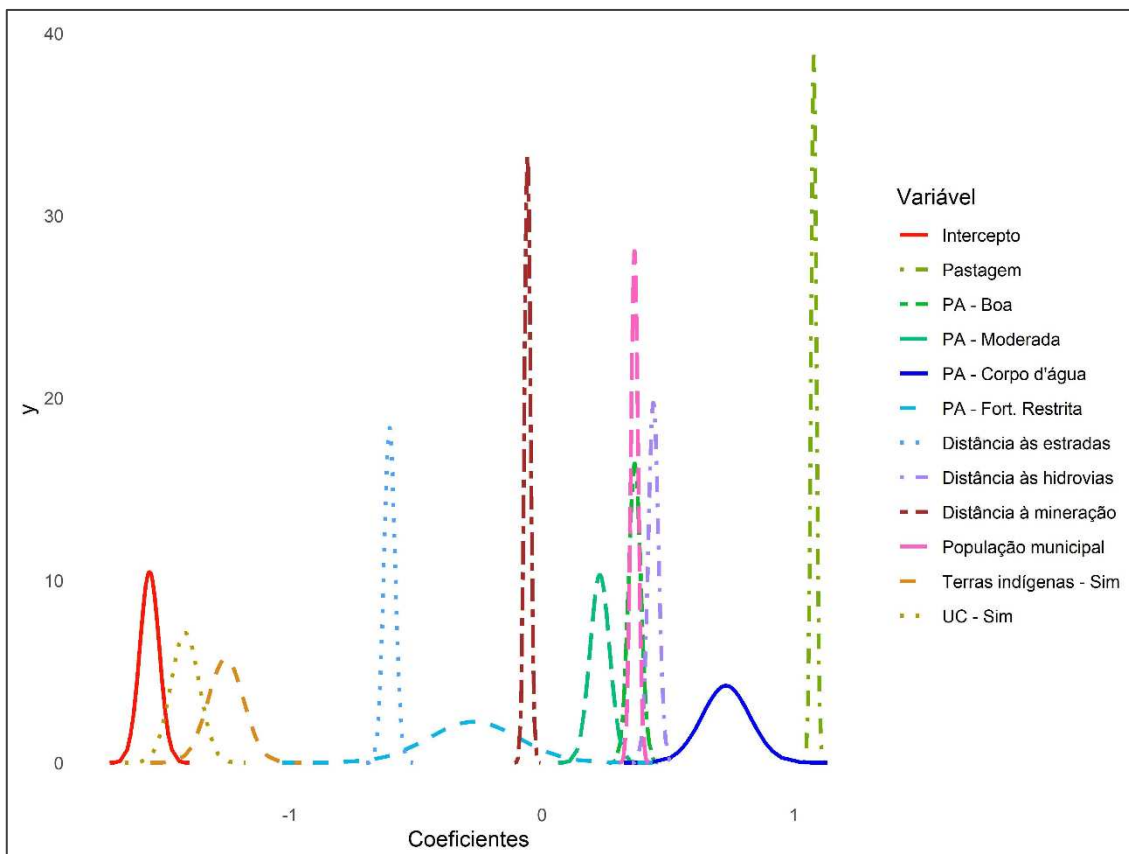


Figura 4 - Distribuições marginais para os efeitos associados às covariáveis no modelo de fragilidade para avaliação da dispersão, posição e associação com relação aos demais coeficientes, calculadas a partir da verossimilhança associada ao modelo Weibull, PA = Potencialidade agrícola, UC = Unidades de conservação.

As distribuições marginais para os hiperparâmetros do modelo podem ser verificadas na Figura 6. É possível observar que o parâmetro alpha da distribuição Weibull possui uma distribuição simétrica com relação ao parâmetro de precisão, utilizado no modelo ICAR. O valor do parâmetro α é de extrema importância pois define a forma da distribuição. Como $\alpha > 1$ para quase todo o suporte da distribuição encontrada, têm-se que a taxa de falha é crescente ao longo do tempo.

A distribuição do parâmetro de precisão é essencial para o modelo, pois indica a variabilidade dos efeitos aleatórios com relação a regiões vizinhas. Ambas as Figuras 6.a

e 6.b, representam distribuições marginais *a posteriori*, após serem utilizados os dados para a atualização dos valores desses parâmetros.

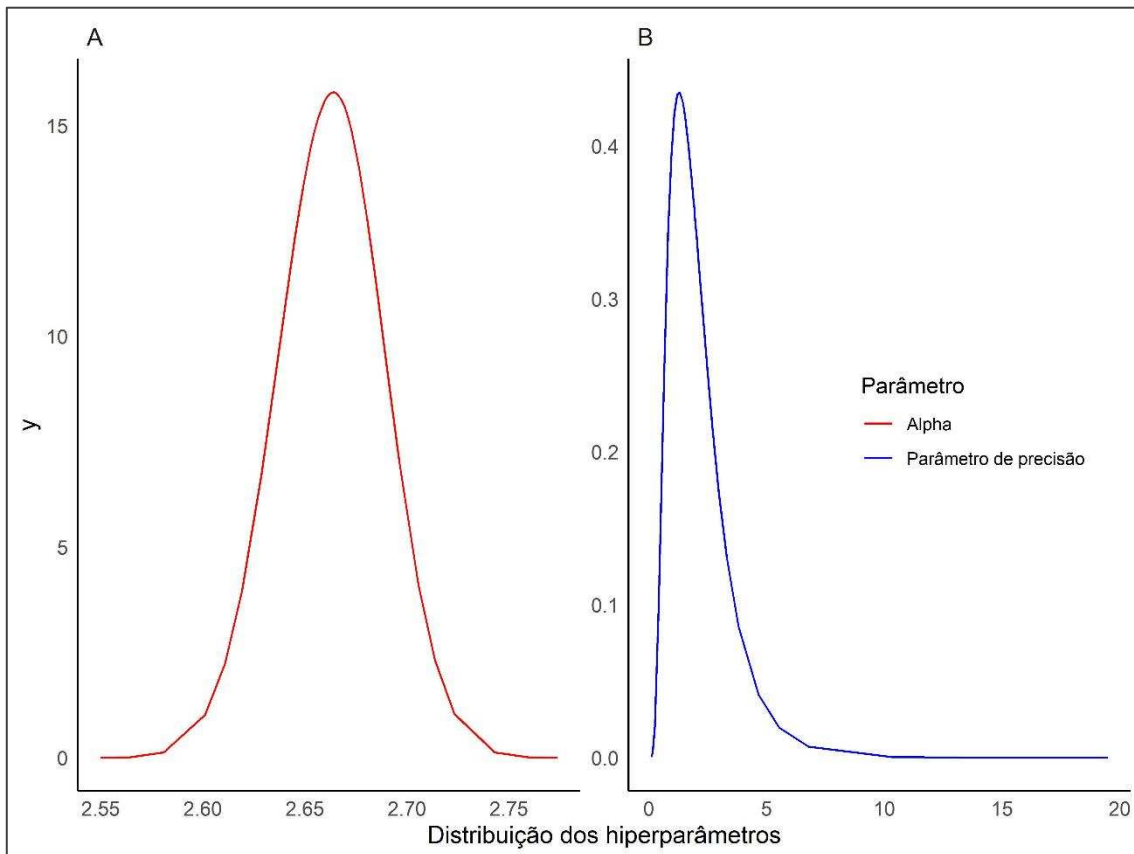


Figura 5 - Distribuições marginais para os hiperparâmetros do modelo de fragilidade associado à distribuição Weibull para a verossimilhança e o modelo autorregressivo ICAR para modelagem do termo de efeito aleatório.

2.4.4. Efeitos aleatórios

Os valores para os efeitos aleatórios observados para cada grupo por meio do modelo de fragilidade estão descritos na Tabela 6.

Tabela 6 - Tabela para identificação dos efeitos aleatórios associados aos grupos identificados por meio do método K-Means para o modelo de fragilidade.

Grupos	w	$\exp(-w.\sigma)$	sd	IC _{2,5%}	IC _{97,5%}	KLD
1	0,9466	0,7006	0,0300	0,6825	0,7137	0
2	-0,1442	1,0557	0,0524	1,0159	1,0984	0
3	-0,7429	1,3221	0,0424	1,2850	1,3688	0
4	-0,2696	1,1067	0,0422	1,0735	1,1433	0
5	0,2102	0,9240	0,0331	0,9007	0,9463	0

w = termo de efeito aleatório associado a cada grupo; σ = scale; KLD = índice KLD (*Kullback Leibler Divergence*); IC = intervalo de credibilidade.

Fonte: O autor (2025).

A partir da Tabela 6 é possível concluir que os grupos 1 e 5 possuem menor fator de aceleração, independentemente do valor das covariáveis. Esses resultados podem ser visualizados espacialmente na Figura 7.

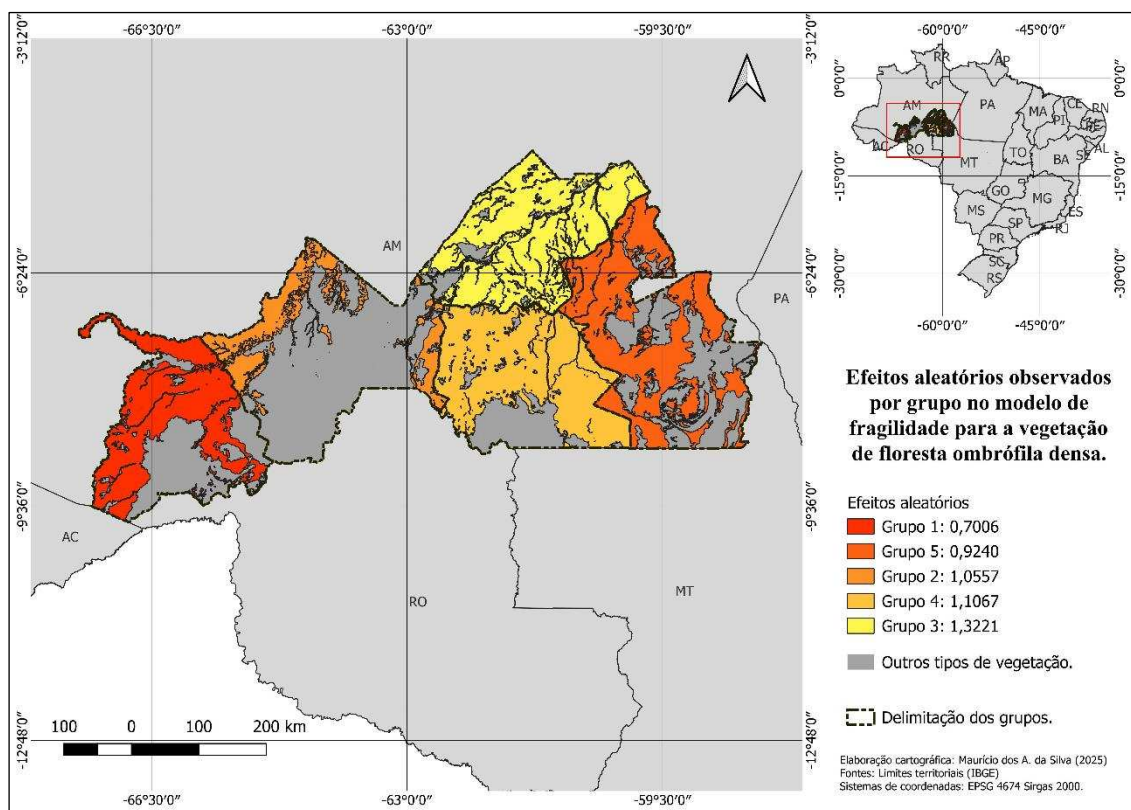


Figura 6 - Efeitos aleatórios para os grupos identificados associados ao desmatamento no modelo de fragilidade para o modelo Weibull.

A delimitação dos efeitos aleatórios não coincide com os grupos, pois estes foram formados utilizando toda a área de estudo, enquanto que o modelo foi ajustado para o tipo de vegetação floresta ombrófila densa. Pela análise da Figura 7, é possível perceber que algumas regiões apresentaram maiores valores para o termo de efeito aleatório, a exemplo do grupo 3. Isso indica que, regiões em que o termo de fragilidade é maior, representam regiões mais frágeis em comparação a outras regiões independentemente dos valores associados às covariáveis e inseridos no modelo. Essas medidas revelam que essas regiões estão mais sujeitas ao desmatamento, e, devido a covariáveis não mensuradas nos dados, apresentam maiores chances de serem desmatadas.

2.5. Discussões

Os resultados obtidos em ambas as abordagens, não paramétrica e o modelo de fragilidade, são coerentes entre si. Isso indica que as conclusões acerca da importância de cada nível das covariáveis são semelhantes, como em unidades de conservação e terras indígenas, em que áreas protegidas possuem um tempo esperado até o desmatamento maior em comparação às áreas que não estão presentes em unidades de conservação e terras indígenas, respectivamente.

A abordagem não paramétrica é adequada em situações em que se deseja analisar individualmente cada covariável com relação ao desmatamento. Já o modelo de fragilidade se mostrou adequado no cenário em que é necessário introduzir o termo de efeito aleatório espacial, para com isso inserir a dependência entre os tempos até que a área não desmatada seja desmatada.

Os resultados descritos na Tabela 5, Figura 2 e Figura 3, fornecem evidências da importância de covariáveis determinantes no desmatamento. Essas covariáveis foram

representadas pela distância do pixel às rodovias, distância às hidrovias, distância a pontos de mineração, pastagem, população, potencialidade agrícola, terras indígenas e unidades de conservação.

Diversos estudos, como os conduzidos por Rojas et al. (2021), Armenteras et al. (2013) e Monjardin-Armenta et al. (2020), mostraram que as covariáveis unidades de conservação, distância até estradas e potencialidade agrícola desempenham um papel significativo para descrever o desmatamento, o que corrobora com os resultados obtidos nesse estudo. Na prática, esse efeito em unidades de conservação pode ser atribuído às políticas de preservação e a todos os movimentos de fiscalização e operações executadas nesses locais. Esse resultado demonstra que o desmatamento nessas regiões está sendo restringido por essas ações, o que demonstra a importância dessas áreas no controle do desmatamento (Soares-Filho et al. 2010).

De acordo com a Tabela 5, o estudo proposto por Rojas et al. (2021) também mostra que a covariável distância até hidrovias possui baixa relevância para descrever o desmatamento na área, o que é coerente com os resultados encontrados nesse estudo. Esse resultado pode ser explicado por diversos motivos. A presença de poucas hidrovias na região de estudo pode ser um fator determinante para a baixa relevância dessa variável no território avaliado, em conjunto ao baixo número de áreas desmatadas próximas a hidrovias. Além disso, pela análise dos intervalos de credibilidade (Tabela 5), é possível ver que o intervalo associado à variável distância até hidrovias contém o valor 1, o que indica a incerteza sobre o parâmetro associado à variável, tornando-a não significativa.

Assim como disposto na resolução CONAMA nº 13/90 (1990), o efeito das áreas de amortecimento no desmatamento é discutido por Jusys (2016a). No estudo presente, as áreas de amortecimento foram adicionadas no intuito de melhorar a modelagem do

desmatamento com a adição de informação a respeito das covariáveis. No entanto, o acréscimo dessas informações não adicionou melhoria relevante ao modelo. Isso pode ser explicado por diversos motivos. Dentre esses, destaca-se a menor área de estudo com relação ao estudo proposto por Jusys (2016a) no estado do Pará, ou mesmo uma fiscalização inadequada nas áreas de amortecimento observadas na região de estudo ao sul do Amazonas.

O tamanho da área de estudo também influencia na significância das covariáveis, uma vez que, quanto maior a área de estudo mais informação se tem a respeito do desmatamento que é explicada por meio dessas covariáveis. Nesse cenário, evidências estatísticas são encontradas a favor do impacto dessas covariáveis no estudo proposto por Armenteras et al. (2013) para o desmatamento na região da Colômbia. No entanto, essa dinâmica é alterada, e algumas deixam de se mostrar importantes para a modelagem do desmatamento ao subdividir o território em 3 grandes regiões e conduzir um estudo para cada região (Armenteras et al. 2013).

As dinâmicas associadas ao desmatamento podem se comportar de maneira semelhante sobre diferentes regiões geoespaciais. Um estudo de simulação foi conduzido por Monjardin-Armenta et al. (2020), em que consideram, a partir de dados anteriores, simular dados para verificar o comportamento de fatores sobre o desmatamento e o reflorestamento. Os dados mostraram que os maiores índices de sensibilidade para o modelo geoespacial aplicado estão relacionados às covariáveis áreas protegidas (Unidades de conservação), distância até estradas e distância até atividades de agricultura e pastagens, o que é coerente com os resultados encontrados nesse estudo, contrastando as diferentes regiões. Monjardin-Armenta et al. (2020), abordam dados da região centro-

oeste do México, enquanto o presente estudo possui enfoque na região centro-sul da Amazônia (Figura 1).

O presente estudo encontrou evidências estatísticas para afirmar que as covariáveis pastagem e distância até pontos de mineração são estatisticamente significativas para descrever o desmatamento na região, e podem ser observadas pela Figura 2 e Figura 3c e na Tabela 5. Um melhor tipo de pastagem, está associada a diversos fatores como a qualidade do solo, a disponibilidade de água e o clima, que favorecem a criação de gado na região. Esse efeito significativo é também esperado e coerente com as dinâmicas do desmatamento observadas, de forma que essas regiões são mais exploradas para as práticas da pecuária.

Por meio da avaliação das distribuições marginais a posteriori (Figura 4), é notório que a variável pastagem possui distribuição marginal concentrada de forma mais densa em torno do seu valor médio, o que acontece também com a variável distância a rodovias. Esse resultado também foi encontrado por Barona et al. (2010).

A variável distância do pixel até pontos de mineração foi significativa para descrever o desmatamento na região (Tabela 5, Figura 3c). Sonter et al. (2017) descrevem como o desmatamento pode ser induzido por atividades de mineração, seja por aumentar a concentração populacional nessas áreas, criação de infraestrutura para essas atividades além do desenvolvimento de cadeias para suprimento de commodities minerais.

Pelos resultados dispostos na Tabela 5 e Figura 2c, as terras indígenas interferem diretamente no tempo esperado até o desmatamento da área correspondente a cada pixel. No estudo proposto por Silva-Junior et al. (2023), fica evidente que áreas dentro de territórios indígenas possuem taxas menores no aumento do índice de desmatamento, o que corrobora também aos resultados obtidos nesse estudo. No entanto, essa importância

é questionada pelo fato de as terras indígenas estarem perdendo força nesse papel de controle do desmatamento (Silva-Junior et al. 2023), uma vez que políticas direcionadas às terras indígenas não têm se mostrado tão eficientes nos últimos anos. O impacto também é atribuído às frequentes atividades de mineração ilegal nessas terras (Silva-Junior et al. 2023).

A partir dos efeitos aleatórios do modelo foi possível verificar a existência de padrões de agrupamento nos valores de fragilidade, indicando grupos de áreas em que o termo associado representa um maior ou menor valor de fragilidade com relação às demais áreas (Hougaard 1995, Klein 2014). Esse fato indica que o tempo até o desmatamento pode ser acrescido ou diminuído com base no valor do termo de efeito aleatório, e com base em efeitos de covariáveis latentes no modelo, isto é, covariáveis que não foram medidas diretamente nos dados.

Adicionalmente, os resultados obtidos com relação aos efeitos aleatórios são coerentes com esses trabalhos propostos para descrever o desmatamento (Tabela 5), de modo que, apesar de serem caracterizados como efeitos aleatórios, a magnitude desses efeitos está visivelmente associada também a essas covariáveis.

Algumas limitações do estudo podem ser elencadas. Dentre essas é possível mencionar o custo computacional para execução do algoritmo, além da implementação para verificação da adequação da distribuição *a posteriori* aos dados. Além disso, é possível destacar a resolução limitada dos dados e a possibilidade de ampliá-la. Para trabalhos futuros, é possível citar o ajuste com os demais tipos de vegetação, a identificação e estudo de outras áreas críticas com fronteira ao arco do desmatamento. O aumento de resolução para os pixels em análise também pode contribuir para uma melhor

descrição do espaço de estudo, haja vista o maior volume de dados e a proximidade com relação às resoluções maiores.

2.6. Conclusões

A análise dos dados por meio do modelo de fragilidade no contexto multivariado se mostrou uma técnica robusta para avaliar o desmatamento. Além disso, é possível concluir que covariáveis como o tipo de pastagens, distância do pixel a estradas, população, unidades de conservação e terras indígenas, se mostraram relevantes na descrição do tempo até o pixel ser desmatado. A observação de regiões mais frágeis em áreas superiores ao conhecido arco do desmatamento, evidencia a tendência do desmatamento para uma expansão cada vez maior nessas áreas.

Referências

- Achcar, J.A.; Martinez, E.Z.; Souza, A.D.P.D.; Tachibana, V.M.; Flores, E.F. 2011. Use of Poisson spatiotemporal regression models for the Brazilian Amazon Forest: malaria count data. *Revista da Sociedade Brasileira de Medicina Tropical* 44: 749–754.
- Armenteras, D.; Cabrera, E.; Rodríguez, N.; Retana, J. 2013. National and regional determinants of tropical deforestation in Colombia. *Regional Environmental Change* 13: 1181–1193.
- Ashine, T.; Muleta, G.; Tadesse, K. 2021. Assessing survival time of heart failure patients: using Bayesian approach. *Journal of Big Data* 8: 156.
- Assunção, R.; Krainski, E. 2009. Neighborhood Dependence in Bayesian Spatial Models. *Biometrical Journal* 51: 851–869.
- Azevedo, D.R.M.; Prates, M.O.; Bandyopadhyay, D. 2020. Alleviating Spatial Confounding in Spatial Frailty Models. .
- Balan, T.A.; Putter, H. 2020. A tutorial on frailty models. *Statistical Methods in Medical Research* 29: 3424–3454.
- Ball, J.G.C.; Petrova, K.; Coomes, D.A.; Flaxman, S. 2022. Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation. *Methods in Ecology and Evolution* 13: 2622–2634.

- Bell, K.P.; Irwin, E.G. 2002. Spatially explicit micro-level modelling of land use change at the rural–urban interface. *Agricultural Economics*.
- Besag, J. 1974. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 36: 192–225.
- Bivand, R.S.; Gómez-Rubio, V. 2021. Spatial survival modelling of business re-opening after Katrina: Survival modelling compared to spatial probit modelling of re-opening within 3, 6 or 12 months. *Statistical Modelling* 21: 137–160.
- Bivand, R.S.; Gómez-Rubio, V.; Rue, H. 2015. Spatial Data Analysis with R - INLA with Some Extensions. *Journal of Statistical Software* 63.
- Butt, E.W.; Baker, J.C.A.; Bezerra, F.G.S.; Von Randow, C.; Aguiar, A.P.D.; Spracklen, D.V. 2023. Amazon deforestation causes strong regional warming. *Proceedings of the National Academy of Sciences* 120: e2309123120.
- Collett, D. 1994. *Modelling Survival Data in Medical Research*. 1. ed. Chapman & Hall, London, UK, 355p.
- Colosimo, E.A.; Giolo, S.R. 2024. *Análise de sobrevivência aplicada*. 2nd ed. Editora Edgard Blucher, São Paulo, SP, 382p.
- Cox, D.R. 1972. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 34: 187–202.
- Dias, F.; Suhadolnik, N.; Camargo, H.; Da Silva, S. 2024. Predicting the pulse of the Amazon: Machine learning insights into deforestation dynamics. *Journal of Environmental Management* 362: 121359.
- Dominguez, D.; Del Villar, L.D.J.; Pantoja, O.; González-Rodríguez, M. 2022. Forecasting Amazon Rain-Forest Deforestation Using a Hybrid Machine Learning Model. *Sustainability* 14: 691.
- Fearnside, P.M. 2005. Deforestation in Brazilian Amazonia: History, Rates, and Consequences. *Conservation Biology* 19: 680–688.
- Fearnside, P.M.; Laurance, W.F. 2012. Infraestrutura na Amazônia: as lições dos planos plurianuais. *Caderno CRH* 25: 87–98.
- Flores, B.M.; Montoya, E.; Sakschewski, B.; Nascimento, N.; Staal, A.; Betts, R.A.; et al. 2024. Critical transitions in the Amazon forest system. *Nature* 626: 555–564.
- Gashu, C.; Aguede, A.E. 2024. Assessing the survival time of women with breast cancer in Northwestern Ethiopia: using the Bayesian approach. *BMC Women's Health* 24: 120.
- Govindarajulu, U.; Bedi, S. 2022. K-means for shared frailty models. *BMC Medical Research Methodology* 22: 11.

- Greenberg, J.A.; Kefauver, S.C.; Stimson, H.C.; Yeaton, C.J.; Ustin, S.L. 2005. Survival analysis of a neotropical rainforest using multitemporal satellite imagery. *Remote Sensing of Environment* 96: 202–211.
- Held, L.; Schrödle, B.; Rue, H. 2010. Posterior and Cross-validators Predictive Checks: A Comparison of MCMC and INLA. In: Kneib, T.; Tutz, G. (Eds.), *Statistical Modelling and Regression Structures*, Physica-Verlag HD, Heidelberg, p.91–110.
- Hojtink, H.; Mulder, J.; Van Lissa, C.; Gu, X. 2019. A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods* 24: 539–556.
- Hougaard, P. 1995. Frailty models for survival data. *Lifetime Data Analysis* 1: 255–273.
- Jusys, T. 2016. Quantifying avoided deforestation in Pará: Protected areas, buffer zones and edge effects. *Journal for Nature Conservation* 33: 10–17.
- Kaplan, E.L.; Meier, P. 1958. Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* 53: 457–481.
- Klein, J.P. 2014. *Handbook of survival analysis*. CRC Press, Boca Raton, Florida, .
- Lavine, M.L.; Hodges, J.S. 2012. On Rigorous Specification of ICAR Models. *The American Statistician* 66: 42–49.
- Miranda, R.D.Q.; Galvêncio, J.D.; Morais, Y.C.B.; Moura, M.S.B.D.; Jones, C.A.; Srinivasan, R. 2018. DRY FOREST DEFORESTATION DYNAMICS IN BRAZIL'S PONTAL BASIN. *Revista Caatinga* 31: 385–395.
- Monteiro, A.M.V.; Rennó, C.D.; Almeida, C.A.; Agônono, E.; Valeriano, D. de M.; Soares, J.V.; et al. 2008. RELATÓRIO TÉCNICO-CIENTÍFICO.
- R Core Team, 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available online: <<https://www.R-project.org/>>.
- Rue, H.; Martino, S.; Chopin, N. 2009. Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 71: 319–392.
- Saha, S.; Saha, M.; Mukherjee, K.; Arabameri, A.; Ngo, P.T.T.; Paul, G.C. 2020. Predicting the deforestation probability using the binary logistic regression, random forest, ensemble rotational forest, REPTree: A case study at the Gumani River Basin, India. *Science of The Total Environment* 730: 139197.
- Sales, V.G.; Strobl, E.; Elliott, R.J.R. 2022. Cloud cover and its impact on Brazil's deforestation satellite monitoring program: Evidence from the cerrado biome of the Brazilian Legal Amazon. *Applied Geography* 140: 102651.
- Schmalz, X.; Biurrun Manresa, J.; Zhang, L. 2023. What is a Bayes factor? *Psychological Methods* 28: 705–718.

Silva, A.C.O.; Fonseca, L.M.G.; Körting, T.S.; Escada, M.I.S. 2020. A spatio-temporal Bayesian Network approach for deforestation prediction in an Amazon rainforest expansion frontier. *Spatial Statistics* 35: 100393.

Silva-Junior, C.H.L.; Silva, F.B.; Arisi, B.M.; Mataveli, G.; Pessôa, A.C.M.; Carvalho, N.S.; et al. 2023. Brazilian Amazon indigenous territories under deforestation pressure. *Scientific Reports* 13: 5851.

Soares-Filho, B.; Moutinho, P.; Nepstad, D.; Anderson, A.; Rodrigues, H.; Garcia, R.; et al. 2010. Role of Brazilian Amazon protected areas in climate change mitigation. *Proceedings of the National Academy of Sciences* 107: 10821–10826.

Souza, C.M.; Z. Shimbo, J.; Rosa, M.R.; Parente, L.L.; A. Alencar, A.; Rudorff, B.F.T.; et al. 2020. Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. *Remote Sensing* 12: 2735.

Vhudzijena, M. 2023. *Modelling mortality heterogeneity using health trajectories and multimorbidity*. UNSW Sydney, (<http://hdl.handle.net/1959.4/101589>).

Wang, X.; Ryan, Y.Y.; Faraway, J.J. 2018. Bayesian Regression Modeling with INLA. .

Zemp, D.C.; Schleussner, C. F.; Barbosa, H.M.J.; Rammig, A. 2017. Deforestation effects on Amazon forest resilience. *Geophysical Research Letters* 44: 6182–6190.

APÊNDICE A - O método K-Means para agrupamento

Nesta seção, serão apresentados os procedimentos para a aplicação do método K-Means para identificação dos grupos que foram utilizados diretamente no modelo de fragilidade.

A aplicação do modelo de fragilidade exige uma estrutura de grupos bem definida, tanto com relação à sua correlação com a variável resposta, como com relação a seu desenho na malha poligonal. Como essa estrutura não foi observada, nem mesmo observado um critério de decisão para identificação da malha de grupos, a partir da malha da variável resposta, foram criados os grupos. Dessa maneira, esse formato automaticamente carrega informação a respeito da variável resposta não contida nos dados.

Foi aplicado o método K-Means para identificação de grupos a partir da dispersão espacial da variável desmatamento, segundo a concentração de pixels desmatados. Esse processo foi estabelecido sobre a configuração da camada raster de pixels desmatados e não desmatados.

Após a reclassificação e poligonalização dos dados de desmatamento, foram salvas somente as feições que indicavam a presença de desmatamento. Para a aplicação do método K-Means foi considerado todo o período de estudo avaliado, de 1987 a 2023.

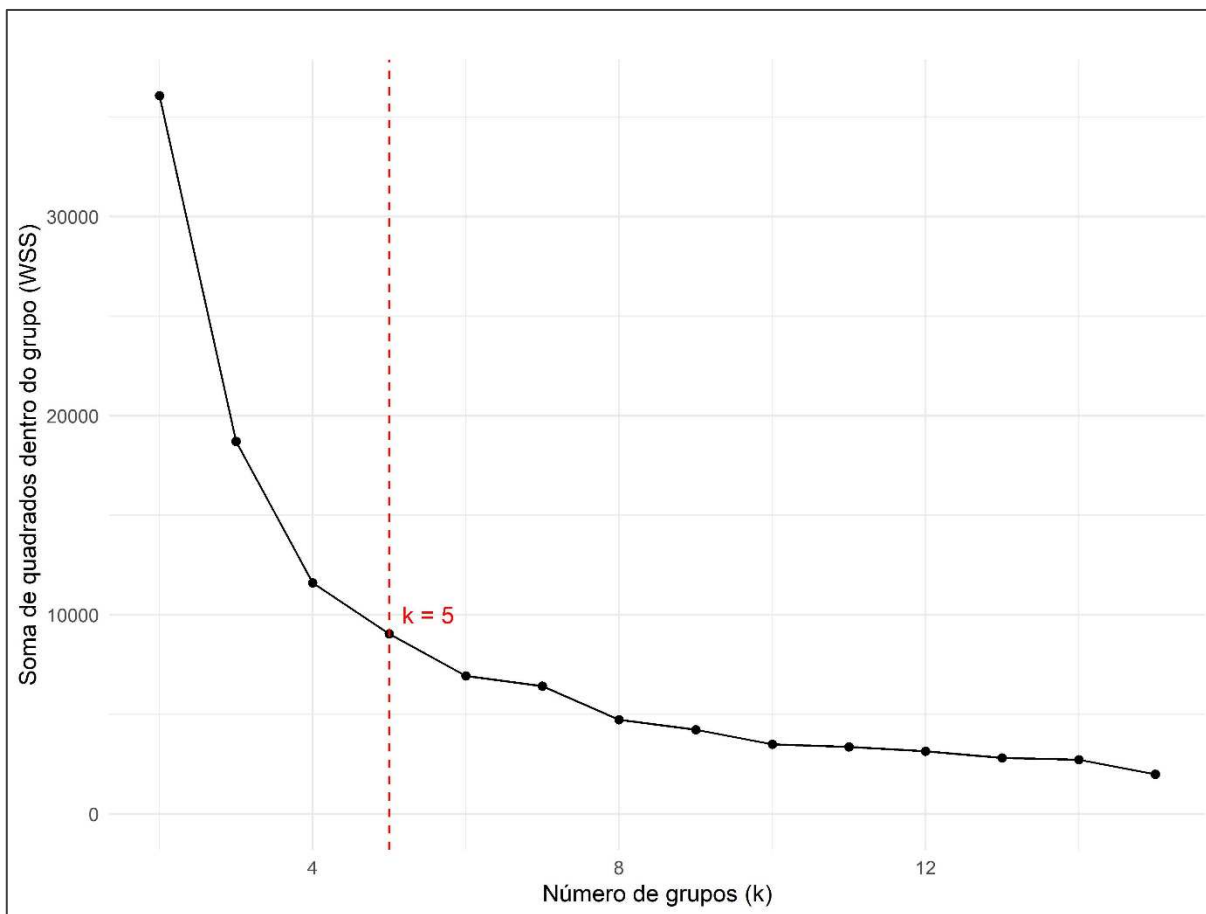


Figura 7 - Gráfico de cotovelo (Elbow Graph) para diferentes valores de k , com representação de 2 a 15 grupos.

A partir da Figura 7 é possível ver que o número de grupos identificado acima, de 5 grupos, se encontra em uma região também considerada ideal no gráfico de cotovelo, uma vez que a partir dele a soma de quadrados dentro do grupo não se altera significativamente.

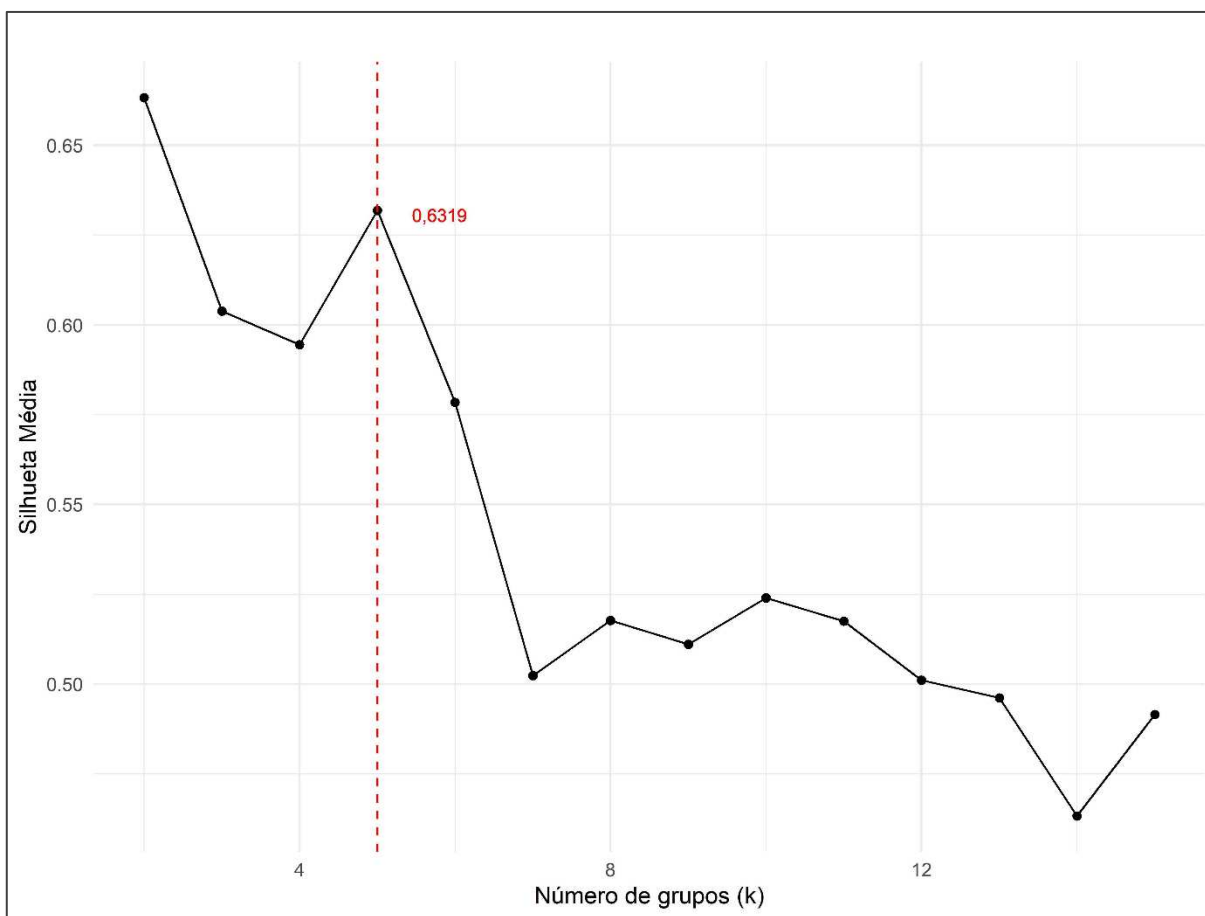


Figura 8 - Silhueta média para diferentes valores de k, com variação para o número de grupos de 2 a 15 grupos.

A partir da Figura 8 é possível perceber que o número de grupos em que a silhueta média é maior acontece quando $K=5$. A partir desse número grupos, os valores de silhueta sempre são menores, portanto, é uma escolha adequada para avaliação. Com essas observações, é possível ver que a quantidade de 5 grupos se torna ideal para a análise da dispersão espacial do desmatamento, uma vez que maximiza a silhueta média e ao mesmo tempo minimiza a soma de quadrados dentro do grupo.

Em contrapartida, apesar da formação dos grupos ter sido executada a partir dos pixels desmatados, por questões de definição do método, os pixels não desmatados também devem ser alocados em algum dos grupos formados, uma vez que essa análise exige que todos os pixels estejam presentes em um grupo. Nesse sentido, para cada pixel não desmatado, foi atribuído a ele o grupo mais próximo definido pelo método K-Means. O mesmo procedimento foi repetido para todos os pixels não desmatados.

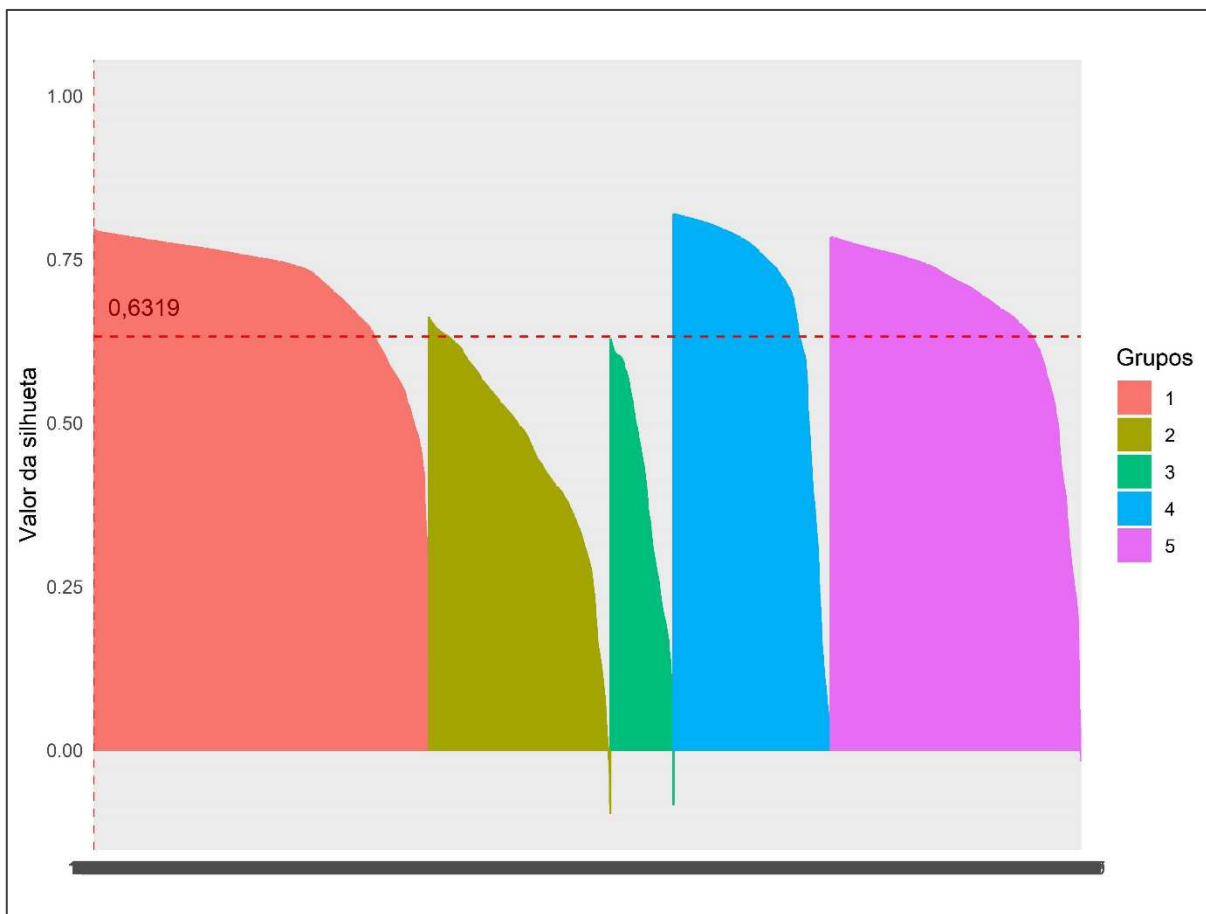


Figura 9 - Gráfico de silhueta para $k = 5$ grupos, com representação individual da silhueta média dentro de cada grupo.

O gráfico de silhueta para $k=5$, apresentado na Figura 9, ilustra como se comportam os valores dentro de cada grupo, permitindo uma visão detalhada sobre consistência interna e a formação dos grupos. Essa análise permite uma melhor compreensão do agrupamento utilizado a partir da classificação geral demonstrada acima.

Dessa forma, os grupos 1, 4, e 5 apresentaram uma maior consistência quanto a sua formação e se mostraram melhor definidos em comparação aos demais pela análise da silhueta, com valores médios superiores a 0,7.

APÊNDICE B – Seleção de variáveis

Nesta seção, serão apresentados os procedimentos referentes à aplicação de uma adaptação do método de Collett para a seleção de covariáveis de forma a manter no modelo somente as covariáveis que se mostram significativas para o ajuste global. Nesse estudo foram considerados o DIC e fator de Bayes para definição e decisão acerca do melhor modelo, tanto com relação ao modelo paramétrico, quanto ao modelo de fragilidade.

Para conduzir o passo 1, as covariáveis foram consideradas individualmente e comparadas com o modelo nulo. Os resultados para essa etapa se encontram na tabela a seguir:

Tabela 7 - Ajustes individuais para verificação do primeiro passo do método de Collett com identificação do DIC e Fator de Bayes como critério de seleção do melhor modelo com relação à cada covariável.

Modelo	Variáveis	Log-Verossimilhança	DIC	Fator de Bayes
Nulo	(Nenhuma)	-24127,64	48242,12	-
1	Unidades de conservação	-21704,94	43384,17	>100
2	Terras indígenas	-23122,54	46219,88	>100
3	Pastagem	-8469,34	16908,22	>100
4	Potencialidade agrícola (1)	-24067,25	48072,99	>100
	Potencialidade agrícola (2)			
	Potencialidade agrícola (3)			
	Potencialidade agrícola (4)			
5	Áreas Urbanas (Densa)	-24099,75	48167,50	>100
	Áreas Urbanas (Pouco Densa)			
6	Distância às rodovias	-18686,27	37345,24	>100
7	Distância à hidrovia	-23651,98	47274,66	>100
8	Distância à mineração	-23762,09	47495,63	>100
9	População	-24136,33	48243,70	0,0002

Fonte: O autor (2025).

Dessa maneira, a variável população foi removida no passo 1 a partir da análise do DIC e Fator de Bayes, indicando evidência a favor do modelo anterior.

No segundo passo, cada variável é removida individualmente e o ajuste comparado com o modelo completo. Os resultados para o passo 2, estão dispostos na Tabela 8.

Tabela 8 - Tabela para verificação do passo 2 no método de Collett com a comparação gerada pela retirada de cada variável com o ajuste global.

Modelo	Variável removida	Log-Verossimilhança	DIC	Fator de Bayes
1	Modelo completo	-7193,78	14217,80	-
2	Distância a mineração	-7263,72	14373,40	>100
3	Áreas Urbanas	-7193,08	14235,61	0,4966
4	Distância a Hidrovias	-7216,27	14277,86	>100
5	Distância às rodovias	-7587,1	15018,99	>100
6	Potencialidade agrícola	-7385,03	14649,84	>100
7	Pastagem	-14437,22	28722,68	>100
8	Terras indígenas	-7358,89	14560,24	>100
9	Unidades de conservação	-7673,73	15190,49	>100

Fonte: O autor (2025).

O terceiro passo foi iniciado com 7 covariáveis, em que a variável população foi removida no passo 1 e a variável áreas urbanas foi removida no passo 2. Portanto o terceiro passo consiste em adicionar a variável população para verificar o comportamento do modelo com relação ao ajuste global. Os resultados para essa verificação se encontram na Tabela 9.

Tabela 9 – Tabela para a verificação do passo 3 da aplicação do método de Collett com representação do DIC e Fator de Bayes para comparar o ajuste global do modelo.

Modelo	Variáveis	Log-Verossimilhança	DIC	Fator de Bayes
1	Modelo Completo	-7193,08	14235,61	-
2	Modelo Completo + População	-6865,23	13564,37	>100

Fonte: O autor (2025).

Por meio do passo 3, é possível ver que a variável população se mostra significativa no ajuste global, sendo necessário, portanto, inseri-la novamente no ajuste do modelo.

O passo 4 consiste em uma nova verificação das covariáveis removidas no passo 2, com a adição da covariável população. Portanto esse passo consiste em comparar novamente o modelo com todas as covariáveis do passo 3 e comparar ao modelo com a adição da covariável áreas urbanas. Os resultados para o passo 4 estão dispostos na Tabela 10.

Tabela 10 - Tabela para verificação do passo 4 da adaptação do método de Collett para validação da remoção das covariáveis no passo 2, com identificação do DIC e Fator de Bayes.

Modelo	Variáveis	Log-Verossimilhança	DIC	Fator de Bayes
1	Modelo Completo	-6865,23	13564,37	-
2	Modelo Completo + Áreas Urbanas	-6868,10	13550,85	0,0567

Fonte: O autor (2025).

É possível ver que novamente a variável áreas urbanas não se mostrou significativa para o modelo, portanto, foi removida do modelo.

O passo 5 consiste então em retirar novamente as covariáveis uma a uma para verificar se todas contribuem globalmente para o modelo de forma significativa. Os resultados para o passo 5 estão dispostos na Tabela 11.

Nesse sentido, após todos os processos executados, a variável áreas urbanas foi removida do modelo. Todas as covariáveis mantidas representaram importância significativa para o ajuste global do modelo.

Tabela 11 - Tabela para verificação do passo 5 para a adaptação do método de Collett para definição das covariáveis selecionadas pelo procedimento, com comparação dos modelos por meio do DIC e Fator de Bayes.

Modelo	Variável removida	Log-Verossimilhança	DIC	Fator de Bayes
1	Modelo completo	-6865,23	13564,37	-
2	Distância à mineração	-6868,01	13585,59	16,12
3	População	-7193,08	14235,61	>100
4	Distância à Hidrovias	-7089,15	14026,98	>100
5	Distância às rodovias	-7368,43	14585,23	>100
6	Potencialidade agrícola	-6969,22	13821,86	>100
7	Pastagem	-13507,46	26867,97	>100
8	Terras indígenas	-7083,11	14012,40	>100
9	Unidades de conservação	-7305,46	14457,64	>100

Fonte: O autor (2025).

APÊNDICE C – Ajuste do modelo paramétrico

Nesta seção, serão apresentados todos os procedimentos considerados no ajuste do modelo paramétrico. Esse processo antecede o ajuste do modelo de fragilidade e foi utilizado para comparação dos resultados obtidos.

Para o ajuste do modelo paramétrico, foram consideradas as distribuições Exponencial, Weibull, Lognormal, Log-Logístico, com o intuito de se obter um modelo que se adequasse minimamente aos dados conforme desejado. A vantagem do modelo paramétrico sobre o modelo não paramétrico é que ele está associado a uma distribuição de probabilidade, sendo, portanto, possível analisar o impacto de cada covariável no modelo global.

Os coeficientes associados aos modelos paramétricos verificados estão dispostos na Tabela 12:

Tabela 12 - Coeficientes estimados para os modelos paramétricos de sobrevivência, ajustados para a modelagem do desmatamento.

Parâmetro	Exponencial		Weibull		Log-normal		Log-logístico	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
β_0	3,1769	0,0281	0,5906	0,0190	0,8173	0,0190	0,4853	0,0175
β_1	1,3714	0,0563	0,5372	0,0219	0,4568	0,0193	0,5052	0,0204
β_2	1,2056	0,0704	0,4756	0,0271	0,4470	0,0244	0,4646	0,0258
β_3	-1,0609	0,0105	-0,4091	0,0056	-0,5144	0,0060	-0,4152	0,0054
β_4	-0,7243	0,0948	-0,2783	0,0361	-0,4602	0,0376	-0,1612	0,0109
β_5	-0,3207	0,0244	-0,1394	0,0094	-0,1677	0,0131	-0,1374	0,0170
β_6	-0,2297	0,0389	-0,0871	0,0148	-0,2794	0,0182	-0,3466	0,0379
β_7	0,2378	0,1785	0,0972	0,0677	0,0373	0,0757	0,0650	0,0697
β_8	0,5520	0,0215	0,2296	0,0085	0,2459	0,0084	0,2440	0,0083
β_9	-0,3827	0,0201	-0,1675	0,0078	-0,1540	0,0096	-0,1902	0,0086
β_{10}	-0,3139	0,0141	-0,1391	0,0055	-0,1775	0,0073	-0,1703	0,0064
β_{11}	0,0601	0,0121	0,0220	0,0046	0,0280	0,0064	0,0208	0,0053
σ (scale)	1,0000	0,3021	-0,9685	0,0032	-0,2300	0,0194	-1,0769	0,0088
Loglik	-10248	-	-6769	-	-7988	-	-6854	-

Fonte: O autor (2025).

A comparação de modelos foi feita por meio do AIC e BIC. Os modelos que apresentaram menores valores para essas medidas são preferíveis. A tabela abaixo representa uma síntese dos valores do AIC e BIC para cada um dos modelos paramétricos:

Tabela 13 - Tabela para comparação de modelos com apresentação das métricas AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion) para os modelos paramétricos avaliados nos dados de desmatamento.

Modelo	AIC	BIC
Exponencial	20520,96	20641,67
Weibull	13564,38	13695,15
LogNormal	16002,86	16133,63
Log-Logístico	13734,05	13864,82

Fonte: O autor (2025).

É possível perceber que o modelo Weibull apresentou os menores valores para o AIC e o BIC.

O modelo Weibull foi escolhido por apresentar um melhor ajuste na avaliação dos resíduos de Cox-Snell, sendo também o modelo que apresentou menores valores para as métricas de comparação de modelos demonstrada na Tabela 13. A avaliação dos resíduos de Cox-Snell para o modelo Weibull pode ser verificada por meio da figura abaixo:

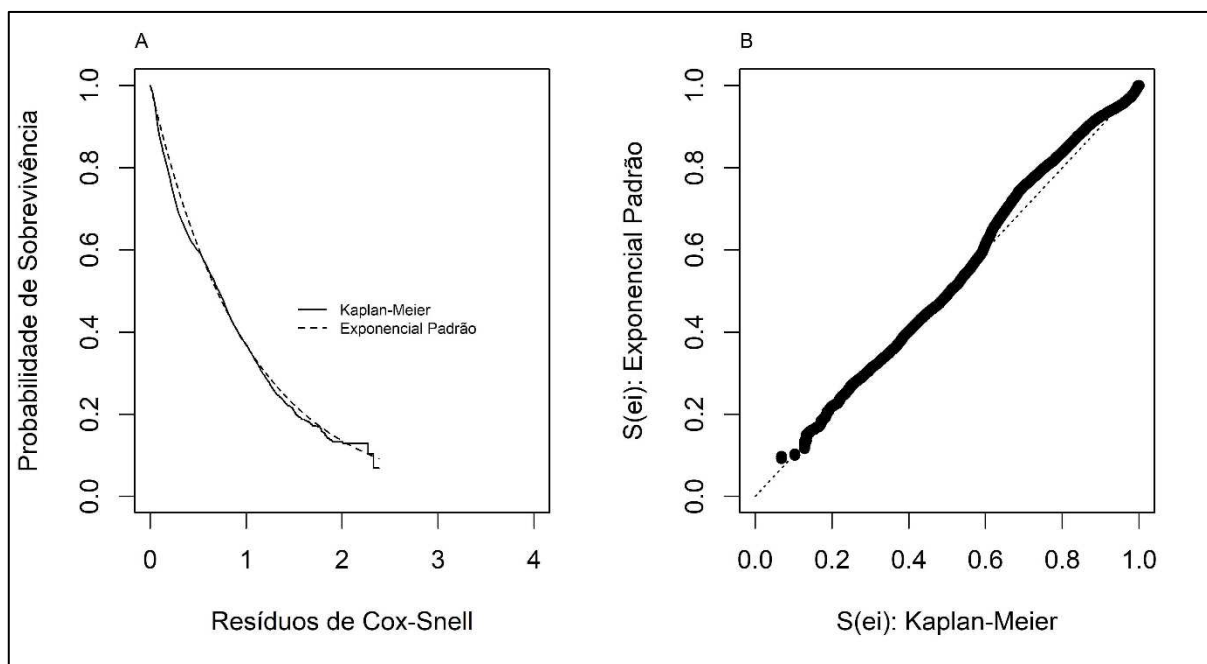


Figura 10 - Resíduos de Cox-Snell para verificação do ajuste do modelo paramétrico Weibull com representação de: (A) comparação do ajuste paramétrico com o estimador não paramétrico de Kaplan-Meier, e (B) comparação dos resíduos obtidos com a distribuição exponencial padrão.

É possível ver que o modelo Weibull se adequou bem aos dados. Com base nessa avaliação, uma análise dos efeitos de cada variável pode ser conduzida. Os coeficientes para o modelo Weibull podem ser verificados por meio da Tabela 14.

De acordo com a Tabela 12 e Tabela 14, para o modelo Weibull, coeficientes dos parâmetros do modelo com sinais negativos, espera-se uma redução no tempo até a área do pixel ser desmatada. Covariáveis como unidades de conservação e terras indígenas apresentam um aumento esperado no tempo até o desmatamento, à medida em que o pixel se encontra nessas áreas, com 71,6% e 61,6% respectivamente. Covariáveis com o coeficiente negativo, como pastagem, apresentam um decaimento esperado no tempo até o desmatamento, com uma redução esperada de 33,67% para essa covariável à medida em que se melhora a qualidade da pastagem em um nível quando mantém-se fixas todas as demais covariáveis.

Tabela 14 - Coeficientes estimados a partir dos dados de desmatamentos para o modelo paramétrico Weibull.

Variável	Coeficiente	Std. Error	z	p valor
Intercepto	0,5906	0,0190	31,15	<0,0001
Unidades de conservação - Sim	0,5372	0,0219	24,49	<0,0001
Terras indígenas - Sim	0,4756	0,0271	17,58	<0,0001
Pastagem	-0,4091	0,0056	-73,33	<0,0001
Potencialidade agrícola - Boa	-0,1394	0,0094	-14,91	<0,0001
Potencialidade agrícola - Moderada	-0,0871	0,0148	-5,89	<0,0001
Potencialidade agrícola - Corpo d'água	-0,2783	0,0361	-7,71	<0,0001
Potencialidade agrícola - Fort. restrita	0,0972	0,0677	1,43	0,15
Distância à rodovias ¹	0,2296	0,0085	26,90	<0,0001
Distância à hidrovias ¹	-0,1675	0,0078	-21,39	<0,0001
População municipal ¹	-0,1391	0,0055	-25,12	<0,0001
Distância à mineração ¹	0,0220	0,0046	4,81	<0,0001

Variável¹ = Variável padronizada

Fonte: O autor (2025).

Na Tabela 15 está representado os coeficientes do modelo paramétrico ajustado pelo *INLA*, no contexto da abordagem bayesiana, sem a adição do termo de fragilidade que caracteriza o efeito aleatório. Esta tabela revela que, os mesmos coeficientes do modelo paramétrico podem ser obtidos após a transformação dos valores de ζ multiplicando-os por σ .

Tabela 15 - Coeficientes do modelo sem o termo de fragilidade, ajustado utilizando a distribuição Weibull para a verossimilhança, com o intervalo de credibilidade com os limites inferiores e superiores a 5% de probabilidade, o valor médio.

Variável	$-\zeta.\sigma$	FA	Sd	IC _{2,5%}	IC _{97,5%}	kld
Intercepto	0,5912	1,8062	0,0381	1,7566	1,8592	0
Unid. de conservação ²	0,5376	1,7119	0,0562	1,6426	1,7861	0
Terras indígenas ³	0,4763	1,6101	0,0703	1,5287	1,6974	0
Pastagem	-0,4091	0,6642	0,0103	0,6589	0,6691	0
PA - Boa	-0,1393	0,8700	0,0244	0,8541	0,8858	0
PA - Moderada	-0,0869	0,9168	0,0389	0,8905	0,9437	0
PA- Corpo d'água	-0,2768	0,7582	0,0948	0,7063	0,8135	0
PA - Fortemente restrita	0,1031	1,1086	0,1783	0,9707	1,2662	0
Distância à rodovias ¹	0,2297	1,2582	0,0218	1,2383	1,2792	0
Distância à hidrovias ¹	-0,1675	0,8458	0,0203	0,8330	0,8585	0
Distância à mineração ¹	0,0220	1,0222	0,0120	1,0132	1,0315	0
População municipal ¹	-0,1391	0,8701	0,0143	0,8608	0,8793	0

¹ Variáveis padronizadas; ζ = Coeficientes do modelo de fragilidade; σ = Scale; $-\zeta.\sigma = \beta$ (Coeficiente encontrado com o ajuste do modelo paramétrico); PA = Potencialidade Agrícola;² Áreas dentro de unidades de conservação; ³Áreas dentro de terras indígenas; FA = Fator de aceleração.

Fonte: O autor (2025).

APÊNDICE D – Script para aplicação do método K-means

```

# K-MEANS

library(sf)
library(ggplot2)
library(dplyr)
library(cluster) # Para calcular a silhueta
library(factoextra) # Para visualização de silhuetas
library(readxl)

pixel_kmeans <- read_excel("Apenas Desmatamento.xlsx",
                           sheet = "Planilha2")

# 1. Carregar a camada do arquivo GeoPackage
shape <- st_read("1990.gpkg")
shape$Pixels_Kmeans = pixel_kmeans$Pixels_kmeans
shape$X1990 = NULL

shape <- subset(shape, shape$Pixels_Kmeans == 1)

# 2. Extrair os centroides dos polígonos
centroids <- st_centroid(shape)

# 3. Extrair coordenadas x e y
coords <- st_coordinates(centroids)

# 4. Amostragem (se necessário)
set.seed(123)
# sample_size <- min(1000, nrow(coords)) # Limitar a 1000 pontos
#ou o número total disponível
# coords_sampled <- coords[sample(nrow(coords), sample_size), ]

# 5. Variar o número de clusters de 2 a 10 e calcular a silhueta
silhouette_scores <- data.frame(k = integer(), silhouette = numeric())

for (k in 2:15) {
  set.seed(123)
  kmeans_result <- kmeans(coords, centers = k)

  # Calcular a silhueta
  silhouette <- silhouette(kmeans_result$cluster, dist(coords))
  avg_silhouette <- mean(silhouette[, 3]) # Média das silhuetas
  silhouette_scores <- rbind(silhouette_scores,
                             data.frame(k = k,
                                         silhouette = avg_silhouette))
}

#grafico de silhueta
# 6. Plotar os resultados
# Obtém o valor da silhueta para k = 5
valor_silhueta <- silhouette_scores$silhouette[silhouette_scores$k == 5]

ggplot(silhouette_scores, aes(x = k, y = silhouette)) +
  geom_line() +
  geom_point() +
  geom_hline(yintercept = valor_silhueta, linetype = "dotted",
            color = "red") + # Linha horizontal pontilhada
  # Linha vertical pontilhada

```

```

geom_vline(xintercept = 5, linetype = "dotted", color = "red") +
geom_text(aes(x = 5.5, y = valor_silhueta + 0.01,
              # Valor da silhueta acima da linha
              label = round(valor_silhueta, 2)), color = "red", size = 4)
+
labs(title = "Silhueta média por número de clusters",
      x = "Número de Clusters (K)",
      y = "Silhueta Média") +
# Mostra todos os valores de 'k' no eixo X
scale_x_continuous(breaks = silhouette_scores$k) +
theme_minimal()

# Converter para data.frame
sil_df <- as.data.frame(silhouette[, 1:3]) %>%
  mutate(cluster = as.factor(cluster),
          observation = row_number())

for (k in 2:6) {
  set.seed(123)
  kmeans_result <- kmeans(coords, centers = k)

  # Calcular a silhueta
  silhouette <- silhouette(kmeans_result$cluster, dist(coords))
  avg_silhouette <- mean(silhouette[, 3]) # Média das silhuetas
  silhouette_scores <- rbind(silhouette_scores,
                             data.frame(k = k, silhouette =
avg_silhouette))
}

fviz_silhouette(silhouette, vertical = TRUE) +
  geom_hline(yintercept = mean(silhouette[, 3]), linetype = "dashed",
            color = "red", linewidth = 1.2) + # Linha horizontal mais
grossa
  labs(title = "Gráfico de Silhueta",
        x = "Valores do Coeficiente de Silhueta",
        y = "Cluster / Label",
        subtitle = "Largura média da silhueta= 0.6319")

table(silhouette)

#Gráfico de cotovelo
wss <- numeric(14) # Para armazenar os valores de WSS para k = 2 a 10

for (k in 2:15) {
  set.seed(123)
  kmeans_result <- kmeans(coords, centers = k)
  # Soma dos quadrados dentro do cluster
  wss[k - 1] <- kmeans_result$tot.withinss
}

# 5. Criar um data frame para plotar
elbow_data <- data.frame(k = 2:15, WSS = wss)

# 6. Plotar o gráfico de cotovelo
ggplot(elbow_data, aes(x = k, y = WSS)) +
  geom_line() +
  geom_point() +
  labs(title = "Gráfico de Cotovelo (Elbow Plot)",
        x = "Número de Clusters (k)",
        y = "Soma dos Quadrados Dentro do Cluster (WSS)") +
  theme_minimal()

```

```

#Adicionando numero do CLUSTER ID
# Escolher o número de clusters (neste caso, 4)
set.seed(123)
kmeans_result <- kmeans(coords, centers = 5)

# Adicionar o cluster resultante à base de dados 'shape'
shape$cluster_id <- kmeans_result$cluster

# Verificar os primeiros resultados
head(shape)

# Salvar o shapefile atualizado em um novo arquivo GeoPackage
st_write(d, "exemplo.gpkg")

# Carregar a camada com pixels desmatados e IDs dos clusters
camada_desmatados <- st_read("shape_clustered.gpkg")

# Carregar a camada com todos os pixels
camada_todos_pixels <- st_read("1990.gpkg")

# Encontrar o índice do pixel desmatado mais próximo para
#cada pixel em 'camada_todos_pixels'
nearest_indices <- st_nearest_feature(camada_todos_pixels,
camada_desmatados)

# Atribuir o ID do cluster do vizinho mais próximo aos pixels não
desmatados
# Assumindo que o campo com o ID dos clusters é
#chamado "cluster_id" na camada desmatados
camada_todos_pixels$cluster_id <-
camada_desmatados$cluster_id[nearest_indices]

# Verificar a nova camada com os IDs dos clusters atribuídos
head(camada_todos_pixels)

# Dissolver diretamente pelos clusters
dissolved_clusters <- aggregate(camada_todos_pixels,
                                by = list(camada_todos_pixels$cluster_id),
                                FUN = function(x) x[1])

#Validando as geometrias
valid_geometries <- st_is_valid(dissolved_clusters)

sum(!valid_geometries)

if (sum(!valid_geometries) > 0) {
  dissolved_clusters <- st_make_valid(dissolved_clusters)
}

# # Verificando os resultados
# plot(dissolved_clusters)
#
# st_write(dissolved_clusters["Group.1"], "exemplo2.gpkg")

# Verificar se todas as geometrias são válidas
valid_geometries <- st_is_valid(dissolved_clusters)

# Contar quantas geometrias são inválidas
num_invalid <- sum(!valid_geometries)

```

```

# Se houver geometrias inválidas, corrigir
if (num_invalid > 0) {
  dissolved_clusters <- st_make_valid(dissolved_clusters)

  # Verificar novamente
  valid_geometries <- st_is_valid(dissolved_clusters)
  invalid_indices <- which(!valid_geometries)

  # Se ainda houver geometrias inválidas, aplicar buffer para corrigir
  if (length(invalid_indices) > 0) {
    dissolved_clusters[invalid_indices, ] <-
      st_buffer(dissolved_clusters[invalid_indices, ], dist = 0.001)

    # Verificar novamente
    valid_geometries <- st_is_valid(dissolved_clusters)
    invalid_indices <- which(!valid_geometries)

    # Corrigir manualmente se necessário
    if (length(invalid_indices) > 0) {
      dissolved_clusters[invalid_indices, ] <-
        st_make_valid(dissolved_clusters[invalid_indices, ])
    }
  }
}

# Visualizar geometrias inválidas restantes
if (length(invalid_indices) > 0) {
  invalid_geometries <- dissolved_clusters[invalid_indices, ]
  plot(invalid_geometries, col = "red", main = "Geometrias Inválidas")
}

# Dissolver novamente se necessário
dissolved_clusters <- dissolved_clusters %>%
  group_by(cluster_id) %>%
  summarize(geometry = st_union(geometry))

# Salvar a nova camada
st_write(dissolved_clusters, "clusters_dissolvidos_corrigidos.gpkg",
  delete_layer = TRUE)

# Visualizar a nova camada
plot(dissolved_clusters, main = "Clusters Dissolvidos Corrigidos")

```

APÊNDICE E – Script para a aplicação do método não paramétrico

```

# Estimador de KAPLAN-MEIER

rm(list=ls())
ls()

library(readr)
library(dplyr)
library(readxl)
library(patchwork)
library(gridExtra)
library(gridGraphics)
library(survival)
library(extrafont)

#Lendo a base de dados
dados = read_excel("Base_Final.xlsx")

#Filtrando a base de dados para pegar somente o tipo de vegetação
#ombrofila densa
dados = dplyr::filter(dados, dados$Vegetação == "Floresta Ombrófila Densa")
table(dados$Vegetação)

#####

dados = subset(dados, dados$Tempo != 0)

#Reclassificando potencialidade agricola
dados$Potencialidade.agricola = ifelse(
  dados$Potencialidade.agricola == "A1", 5,
  ifelse(dados$Potencialidade.agricola == "A2", 4,
    ifelse(dados$Potencialidade.agricola == "B", 3,
      ifelse(dados$Potencialidade.agricola == "C", 0,
        ifelse(dados$Potencialidade.agricola ==
          "Corpo d'água continental", 2, "-
      ))))
)
table(dados$Potencialidade.agricola)

#Reclassificando as variáveis
dados$t = as.numeric(dados$Tempo)
dados$Pedologia = as.factor(dados$Pedologia)
dados$`Potencialidade agricola` = as.factor(dados$Potencialidade.agricola)
dados$`Trecho hidroviário` = as.factor(dados$Hidrovia)
dados$`Trecho rodoviário` = as.factor(dados$Rodovia)
dados$`Terras indígenas` = as.factor(dados$terras_indigenas)
dados$`Áreas urbanizadas` = as.factor(dados$Areas_urbanas)
dados$`Unidades de conservação` = as.factor(dados$Unidades_de_conservacao)
dados$Pastagem = as.numeric(dados$Pastagem)
dados$Mineracao = as.factor(dados$Mineracao)
dados$e = dados$Censura

#Removendo o tempo=0
dados = subset(dados, dados$t != 0)

# Formatando as variáveis
dados$Unidades_de_conservacao[is.na(dados$Unidades_de_conservacao)] <- 0
dados$terras_indigenas[is.na(dados$terras_indigenas)] = 0

```

```

dados$`Areas_urbanas`[is.na(dados$`Areas_urbanas`)] = 0
table(dados$`Areas_urbanas`)

# Definir o layout para 3 linhas e 3 colunas, mas com um espaço em branco
#na última posição

# Organizando os gráficos
tiff("(TIFF) Não paramétrico Fig1.tiff",width=200,height=150,units="mm",
     res=600,family='Calibri')

layout.m <- matrix(c(rep(1:3, each=2),
                    0, rep(4:5, each=2), 0),
                  nrow=2, byrow=T)

layout(layout.m)
# Ajustando as margens para garantir espaço para os títulos
par(mar=c(4, 4.5, 3.5, 0.5)) # Aumentando a margem superior (3)

#ESTIMADOR DE KAPLAN-MEIER
attach(dados)

#Unidades de conservação
require(survival)

#Comparacao das duas especies
ekm<- survfit(Surv(dados$t,dados$Censura) ~ dados$Unidades_de_conservacao,
              data=dados)
summary(ekm)
print(ekm, print.rmean=TRUE)
quantile(ekm, probs = c(0.90,0.8,0.75,0.7,0.6, 0.5,0.4,0.3,0.2, 0.10),
         conf.int = TRUE)
#fun = event, cumhaz ou retemorar fun para ter a sobrevivencia
g1 = plot(ekm, lwd=1.6,col=c("blue","red"), xlab="Tempo (anos)",
         ylab = "S(t) estimada", xlim=c(0,37), ylim=c(0.91, 1),
         main="")
axis(1, at=seq(0, 37, by=5))

# Adicionando o título na parte superior esquerda
mtext("A", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

legend(0.0, 0.95, col=c("red", "blue"),
       legend=c(expression(italic("Protegida")),
                expression(italic("Não protegida"))),
       lwd=2, bty="n", cex=0.8)
text(0.0, 0.951, "Área", pos=4, cex=0.8, font=2)
text(0.0, 0.914, "p-valor: <0,0001", pos=4, cex=0.8, font=2)

#Teste logrank para comparacao das curvas
survdif(Surv(t,Censura)~ dados$Unidades_de_conservacao,rho=0)
#Teste Wilcoxon para comparacao das curvas
survdif(Surv(t,Censura)~ dados$Unidades_de_conservacao,rho=1)

#####

#Áreas urbanizadas
ekm<- survfit(Surv(t,dados$Censura)~ dados$`Áreas urbanizadas`,data=dados)
summary(ekm)
print(ekm, print.rmean=TRUE)
quantile(ekm, probs = c(0.90,0.8,0.75,0.7,0.6, 0.5,0.4,0.3,0.2, 0.10),
         conf.int = TRUE)
#fun = event, cumhaz ou retemorar fun para ter a sobrevivencia

```

```

g2 = plot(ekm, lwd=1.6, col=c("red", "blue"), xlab="Tempo (anos)",
         ylab="S(t) estimada", xlim=c(0,37), ylim=c(0.2, 1),
         main="")
axis(1, at=seq(0, 37, by=5))

# Adicionando o título na parte superior esquerda
mtext("B", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

legend(0.0, 0.57, col=c("red", "blue"),
       legend=c(expression(italic("Densa")),
                expression(italic("Pouco densa"))),
       lwd=2, bty="n", cex=0.8) # Aqui o 'cex' diminui o tamanho da letra
text(0.0, 0.58, "Áreas urbanas", pos=4, cex=0.8, font=2)
text(0.0, 0.23, "p-valor: 0,01", pos=4, cex=0.8, font=2)
#
#Teste logrank para comparacao das curvas
survdif(Surv(t,e)~ dados$`Áreas urbanizadas`,rho=0)
#Teste Wilcoxon para comparacao das curvas
survdif(Surv(t,e)~ dados$`Áreas urbanizadas`, data=dados,rho=1)

#####

#Terras indígenas
ekm<- survfit(Surv(t,e)~ dados$terras_indigenas ,data=dados)
summary(ekm)
print(ekm, print.rmean=TRUE)
quantile(ekm, probs = c(0.90,0.8,0.75,0.7,0.6, 0.5,0.4,0.3,0.2, 0.10),
         conf.int = TRUE)
#fun = event, cumhaz ou retemporar fun para ter a sobrevivencia
g3 = plot(ekm, lwd=1.6, lty=c(1,1), col=c("red", "blue"), xlab="Tempo
(anos)",
         ylab="S(t) estimada", xlim=c(0,37), ylim=c(0.88, 1),
         main="")
axis(1, at=seq(0, 37, by=5))

# Adicionando o título na parte superior esquerda
mtext("C", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

legend(0.0, 0.935, lty=c(1,1), col=c("red", "blue"),
       legend=c(expression(italic("Não")),
                expression(italic("Sim"))),
       lwd=2, bty="n", cex=0.8) # Aqui o 'cex' diminui o tamanho da letra
text(0.0, 0.936, "Terras indígenas", pos=4, cex=0.8, font=2)
text(0.0, 0.885, "p-valor: <0,0001", pos=4, cex=0.8, font=2)
#
#Teste logrank para comparacao das curvas
survdif(Surv(t,Censura)~ dados$terras_indigenas,rho=0)
#Teste Wilcoxon para comparacao das curvas
survdif(Surv(t,e)~ dados$`Terras indígenas`, data=dados,rho=1)

#####

table(dados$`Potencialidade agrícola`)

#Potencialidade agricola
ekm<- survfit(Surv(t,e)~ as.factor(dados$`Potencialidade agrícola`) ,
             data=dados)
summary(ekm)
print(ekm, print.rmean=TRUE)
quantile(ekm, probs = c(0.90,0.8,0.75,0.7,0.6, 0.5,0.4,0.3,0.2, 0.10),
         conf.int = TRUE)

```

```

#fun = event, cumhaz ou retemorar fun para ter a sobrevivencia
g4 = plot(ekm, lwd=1.6, col=c("blue", "red", "black", "orange", "purple"),
          xlab="Tempo (anos)",
          ylab="S(t) estimada", xlim=c(0,37), ylim=c(0.94, 1),
          main="")
axis(1, at=seq(0, 37, by=5))

# Adicionando o título na parte superior esquerda
mtext("D", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

legend(-0.8, 0.968, col=c("blue", "red", "black", "orange", "purple"),
       legend=c(expression(italic("Corpo d'água")),
                 expression(italic("Fortemente restrita")),
                 expression(italic("Restrita")),
                 expression(italic("Moderada")),
                 expression(italic("Boa")))
       ),
       lwd=2, bty="n", cex=0.8, y.intersp = 0.8)

text(-0.8, 0.969, "Potencialidade agrícola", pos=4, cex=0.8, font=2)
text(0.0, 0.942, "p-valor: <0,0001", pos=4, cex=0.8, font=2)

#Teste logrank para comparacao das curvas
survdiff(Surv(t,e)~ dados$`Potencialidade agrícola`,rho=0)
#Teste Wilcoxon para comparacao das curvas
survdiff(Surv(t,e)~ dados$`Potencialidade agrícola`, data=dados,rho=1)

#####

table(dados$Pastagem)

#Potencialidade agricola
ekm<- survfit(Surv(t,e)~ as.factor(dados$Pastagem) ,data=dados)
summary(ekm)
print(ekm, print.rmean=TRUE)
quantile(ekm, probs = c(0.90,0.8,0.75,0.7,0.6, 0.5,0.4,0.3,0.2, 0.10),
         conf.int = TRUE)

#fun = event, cumhaz ou retemorar fun para ter a sobrevivencia
g5 = plot(ekm, lwd=1.6, col=c("red", "blue", "black"), xlab="Tempo (anos)",
          ylab="S(t) estimada", xlim=c(0,37), ylim=c(0.3, 1),
          main="")

# Adicionando o título na parte superior esquerda
mtext("E", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

axis(1, at=seq(0, 37, by=5))
legend(0.0, 0.64, col=c("red", "blue", "black"),
       legend=c(expression(italic("Baixa")),
                 expression(italic("Média")),
                 expression(italic("Alta")))
       ),
       lwd=2, bty="n", cex=0.8) # Aqui o 'cex' diminui o tamanho da letra
# Adiciona um título acima da legenda
text(0.0, 0.65, "Qualidade da pastagem", pos=4, cex=0.8, font=2)
text(0.0, 0.32, "p-valor: <0,0001", pos=4, cex=0.8, font=2)
#
#Teste logrank para comparacao das curvas
survdiff(Surv(t,e)~ dados$Pastagem,rho=0)
#Teste Wilcoxon para comparacao das curvas
survdiff(Surv(t,e)~ dados$`Potencialidade agrícola`, data=dados,rho=1)

```

```

dev.off()

#####

# CATEGORIZANDO AS VARIÁVEIS QUANTITATIVAS

# Criar a variável categorizada
dados$dist_estrada_categorizada <- ifelse(
  dados$distancia_estrada <= median(dados$distancia_estrada),
  "0 a 44", # Categoria para valores abaixo ou iguais à mediana
  "44 a 208,4" # Categoria para valores acima da mediana
)

# Criar a variável categorizada
dados$dist_hidrovia_categorizada <- ifelse(
  dados$distancia_hidrovia <= median(dados$distancia_hidrovia),
  "0 a 81,7", # Categoria para valores abaixo ou iguais à mediana
  "81,7 a 500" # Categoria para valores acima da mediana
)

# Criar a variável categorizada
dados$população_categorizada <- ifelse(
  dados$População <= median(dados$População),
  "0 a 47685", # Categoria para valores abaixo ou iguais à mediana
  "47685 a 57405" # Categoria para valores acima da mediana
)

# Criar a variável categorizada
dados$dist_mineracao_categorizada <- ifelse(
  dados$distancia_mineracao <= median(dados$distancia_mineracao),
  "0 a 194,6", # Categoria para valores abaixo ou iguais à mediana
  "194,6 a 391,1" # Categoria para valores acima da mediana
)

#####

tiff("(TIFF) Não paramétrico Fig2.tiff",width=200,height=180,units="mm",
     res=600,family='Calibri')
# CRIANDO OS GRAFICOS PARA AS VARIÁVEIS QUANTITATIVAS
par(mfrow = c(2,2))
#DISTANCIA A ESTRADA
ekm<- survfit(Surv(dados$t,dados$Censura) ~
dados$dist_estrada_categorizada,
             data=dados)
summary(ekm)
print(ekm, print.rmean=TRUE)
quantile(ekm, probs = c(0.90,0.8,0.75,0.7,0.6, 0.5,0.4,0.3,0.2, 0.10),
        conf.int = TRUE)
#fun = event, cumhaz ou retemorar fun para ter a sobrevivencia
g1 = plot(ekm, lwd=1.6,col=c("blue","red"), xlab="Tempo (anos)",
        ylab = "S(t) estimada", xlim=c(0,37), ylim=c(0.90, 1),
        main="")
axis(1, at=seq(0, 37, by=5))

# Adicionando o título na parte superior esquerda
mtext("A", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

legend(0.0, 0.95, col=c("blue", "red"),
       legend=c(expression(italic("0 a 44")),
               expression(italic("44 a 208,4"))),
       lwd=2, bty="n", cex=0.8)

```

```

text(0.0, 0.951, "Distância (Km)", pos=4, cex=0.8, font=2)
text(0.0, 0.91, "p-valor: <0,0001", pos=4, cex=0.8, font=2)

#####

#DISTANCIA A HIDROVIA
ekm<- survfit(Surv(dados$t,dados$Censura) ~
dados$dist_hidrovia_categorizada,
              data=dados)
summary(ekm)
print(ekm, print.rmean=TRUE)
quantile(ekm, probs = c(0.90,0.8,0.75,0.7,0.6, 0.5,0.4,0.3,0.2, 0.10),
          conf.int = TRUE)
#fun = event, cumhaz ou retemorar fun para ter a sobrevivencia
gl = plot(ekm, lwd=1.6,col=c("blue","red"), xlab="Tempo (anos)",
          ylab = "S(t) estimada", xlim=c(0,37), ylim=c(0.90, 1),
          main="")
axis(1, at=seq(0, 37, by=5))

# Adicionando o título na parte superior esquerda
mtext("B", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

legend(0.0, 0.95, col=c("blue", "red"),
       legend=c(expression(italic("0 a 81,7")),
                 expression(italic("81,7 a 500"))),
       lwd=2, bty="n", cex=0.8)
text(0.0, 0.951, "Distância (Km)", pos=4, cex=0.8, font=2)
text(0.0, 0.91, "p-valor: <0,0001", pos=4, cex=0.8, font=2)

#####

#DISTANCIA A MINERAÇÃO
ekm<- survfit(Surv(dados$t,dados$Censura) ~
dados$dist_mineracao_categorizada,
              data=dados)
summary(ekm)
print(ekm, print.rmean=TRUE)
quantile(ekm, probs = c(0.90,0.8,0.75,0.7,0.6, 0.5,0.4,0.3,0.2, 0.10),
          conf.int = TRUE)
#fun = event, cumhaz ou retemorar fun para ter a sobrevivencia
gl = plot(ekm, lwd=1.6,col=c("blue","red"), xlab="Tempo (anos)",
          ylab = "S(t) estimada", xlim=c(0,37), ylim=c(0.935, 1),
          main="")
axis(1, at=seq(0, 37, by=5))

# Adicionando o título na parte superior esquerda
mtext("C", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

legend(0.0, 0.97, col=c("blue", "red"),
       legend=c(expression(italic("0 a 194,6")),
                 expression(italic("194,6 a 391,1"))),
       lwd=2, bty="n", cex=0.8)
text(0.0, 0.971, "Distância (Km)", pos=4, cex=0.8, font=2)
text(0.0, 0.94, "p-valor: 0,02", pos=4, cex=0.8, font=2)

#####

#DISTANCIA A POPULAÇÃO
ekm<- survfit(Surv(dados$t,dados$Censura) ~ dados$população_categorizada,
              data=dados)
summary(ekm)

```

```

print(ekm, print.rmean=TRUE)
quantile(ekm, probs = c(0.90,0.8,0.75,0.7,0.6, 0.5,0.4,0.3,0.2, 0.10),
        conf.int = TRUE)
#fun = event, cumhaz ou retemorar fun para ter a sobrevivencia
g1 = plot(ekm, lwd=1.6,col=c("blue","red"), xlab="Tempo (anos)",
        ylab = "S(t) estimada", xlim=c(0,37), ylim=c(0.935, 1),
        main="")
axis(1, at=seq(0, 37, by=5))

# Adicionando o título na parte superior esquerda
mtext("D", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

legend(0.0, 0.97, col=c("red","blue"),
       legend=c(expression(italic("0 a 47685")),
               expression(italic("47685 a 57405"))),
       lwd=2, bty="n", cex=0.8)
text(0.0, 0.971, "Habitantes", pos=4, cex=0.8, font=2)
text(0.0, 0.94, "p-valor: <0,0001", pos=4, cex=0.8, font=2)

dev.off()

#####

#Teste logrank para comparacao das curvas
survdif(Surv(t,Censura)~ dados$dist_estrada_categorizada,rho=0)
#Teste Wilcoxon para comparacao das curvas
survdif(Surv(t,Censura)~ dados$dist_estrada_categorizada,rho=1)

#Teste logrank para comparacao das curvas
survdif(Surv(t,Censura)~ dados$dist_hidrovia_categorizada,rho=0)
#Teste Wilcoxon para comparacao das curvas
survdif(Surv(t,Censura)~ dados$dist_hidrovia_categorizada,rho=1)

#Teste logrank para comparacao das curvas
survdif(Surv(t,Censura)~ dados$dist_mineracao_categorizada,rho=0)
#Teste Wilcoxon para comparacao das curvas
survdif(Surv(t,Censura)~ dados$dist_mineracao_categorizada,rho=1)

#Teste logrank para comparacao das curvas
survdif(Surv(t,Censura)~ dados$população_categorizada,rho=0)
#Teste Wilcoxon para comparacao das curvas
survdif(Surv(t,Censura)~ dados$população_categorizada,rho=1)

```

APÊNDICE F – Script para o ajuste do modelo paramétrico

```

# MODELO PARAMÉTRICO

rm(list=ls())
ls()

options(digits = 4, scipen = 999)
library(readxl)
library(readr)
library(dplyr)
library(flexsurv)
require(survival)
library(knitr)

#LENDO OS DADOS
dados<-read_excel("Base_Final.xlsx")
colnames(dados)
str(dados)

#Filtrando a base de dados para pegar somente o tipo de vegetação
#Floresta ombrofila densa
dados = dplyr::filter(dados, dados$Vegetação == "Floresta Ombrófila Densa")

attach(dados)

#Removendo o tempo=0
dados = subset(dados, dados$Tempo!= 0)

#Modelos com Covariáveis

str(dados)

dados$Areas_urbanas = ifelse(is.na(dados$Areas_urbanas), 0,
                             dados$Areas_urbanas)
dados$Unidades_de_conservacao =
ifelse(is.na(dados$Unidades_de_conservacao),
        0,dados$Unidades_de_conservacao)
dados$Pastagem = ifelse(is.na(dados$Pastagem ), 0, dados$Pastagem )
dados$`11. (Buffer) Unidades de conservação` =
  ifelse(is.na(dados$`11. (Buffer) Unidades de conservação`),
        0,dados$`11. (Buffer) Unidades de conservação`)
dados$terras_indigenas = ifelse(is.na(dados$terras_indigenas),
                                0, dados$terras_indigenas)

str(dados)

# 1-Modelo Exponencial
ajust1<-survreg(Surv(Tempo/100, Censura) ~ Unidades_de_conservacao+
                terras_indigenas+Pastagem+ Potencialidade.agrícola +
                Areas_urbanas + scale(distancia_estrada) +
                scale(distancia_hidrovia)+
                scale(distancia_mineração) + scale(População),
                dist = "exponential", data = dados)

summary(ajust1)
table(dados$`Unidades de conservacao`)

```

```

unique(dados$`Terras indigenas`)

xb1<-exp(ajust1$linear.predictors);xb1
ste<-exp(-((dados$Tempo/100)/xb1));ste # Funcao de sobrevivencia
ajust1$loglik
AIC(ajust1)
BIC(ajust1)

#Calculo do AIC (-2*loglik+2*(p+k) em que p é o num de param da dist e k
# é o num. de parametros (sem o intercepto)
ajust1_AIC<- -2*ajust1$loglik[2]+2*(1+7);ajust1_AIC
#Calculo do BIC (-2*loglik+(p+k)*log(r) #r é o tamanho da amostra
r<-length(dados$Censura);r
ajust1_BIC<- -2*ajust1$loglik[2]+(1+7)*log(r);ajust1_BIC

options(scipen = 999)
options(digits = 4)

#Calcular o Std. Error do exp(Log(scale))
vb <- vcov(ajust1);vb #ou ajust2$var
grad <- ajust1$scale;grad #ja em exponencial
vb2 <- vcov(ajust1)[9,9];vb2
vG <- grad %*% vb2 %*% grad;vG
sqrt(vG) # Std. Error do scale

#####

# AJUSTE DO MODELO PARAMÉTRICO WEIBULL

ajust2<-survreg(Surv(dados$Tempo/100, dados$Censura) ~
Unidades_de_conservacao+
                terras_indigenas+Pastagem+ Potencialidade.agrícola +
                scale(distancia_estrada) + scale(distancia_hidrovia)+
                scale(distancia_mineração) + scale(População),
                dist = "weibull",
                data=dados)
summary(ajust2)
round(summary(ajust2)$table, 4)

citation("survival")

predict(ajust2, type = "response")
xb2<-exp(ajust2$linear.predictors);xb2
gamma2<-1/ajust2$scale;gamma2
stw<-exp(-((dados$Tempo/100)/xb2)^gamma2);stw
ajust2$loglik
AIC(ajust2)
BIC(ajust2)
#
#Calculo do AIC (-2*loglik+2*(p+k) em que p é o num de param da
# dist e k é o num de covariaveis
ajust2_AIC<- -2*ajust2$loglik[2]+2*(2+7);ajust2_AIC
#Calculo do BIC (-2*loglik+(p+k)*log(r) #r é o tamanho da amostra
r<-length(dados$Censura);r
ajust2_BIC<- -2*ajust2$loglik[2]+(2+7)*log(r);ajust2_BIC

#Calcular o Std. Error do exp(Log(scale))
vb <- vcov(ajust2);vb #ou ajust2$var
grad <- ajust2$scale;grad #ja em exponencial
vb2 <- vcov(ajust2)[9,9];vb2

```

```

vG <- grad %*% vb2 %*% grad;vG
sqrt(vG) # Std. Error do scale

# 3-Modelo Log-normal (intercepto é o mi e o Scale é o sigma
ajust3<-survreg(Surv(dados$Tempo/100, dados$Censura) ~
Unidades_de_conservacao+
      terras_indigenas+Pastagem+ Potencialidade.agrícola +
      Areas_urbanas + scale(distancia_estrada) +
      scale(distancia_hidrovia)+
      scale(distancia_mineração) + scale(População),
      dist = "lognorm",
      data = dados)
summary(ajust3)
predict(ajust3, type= "response")
ajust3$linear.predictors
AIC
BIC

xb3<-ajust3$linear.predictors;xb3
sigma3<-ajust3$scale;sigma3
stln<-pnorm((-log(dados$Tempo/100)+ xb3)/sigma3);stln
ajust3$loglik
AIC(ajust3)
BIC(ajust3)

unique(dados$Pastagem)
#Calculo do AIC (-2*loglik+2*(p+k) em que p é o num de param da
# dist e k é o num de covariaveis
ajust3_AIC<- -2*ajust3$loglik[2]+2*(2+8);ajust3_AIC
#Calculo do BIC (-2*loglik+(p+k)*log(r) #r é o tamanho da amostra
r<-length(dados$e);r
ajust3_BIC<- -2*ajust3$loglik[2]+(2+9)*log(r);ajust3_BIC

#Calcular o Std. Error do exp(Log(scale))
vb <- vcov(ajust3);vb #ou ajust3$var
grad <- ajust3$scale;grad #ja em exponencial
vb2 <- vcov(ajust3)[3,3];vb2
vG <- grad %*% vb2 %*% grad;vG
sqrt(vG) # Std. Error do scale

# 4-Modelo Log-logistico
ajust4<-survreg(Surv(dados$Tempo/100, dados$Censura) ~
Unidades_de_conservacao+
      terras_indigenas+Pastagem+ Potencialidade.agrícola +
      Areas_urbanas + scale(distancia_estrada) +
      scale(distancia_hidrovia)+
      scale(distancia_mineração) + scale(População),
      dist = "loglogistic", data=dados)
summary(ajust4)
kable(d$table)
ajust4
e = predict(ajust4, type = "response")

xb4<-exp(ajust4$linear.predictors);xb4
gamma4<-1/ajust4$scale;gamma4
stllog<- 1/(1+(dados$Tempo/xb4)^gamma4);stllog
ajust4$loglik
AIC(ajust4)
BIC(ajust4)

#Calcular o Std. Error do exp(Log(scale))

```

```

vb <- vcov(ajust4);vb #ou ajust3$var
grad <- ajust4$scale;grad #ja em exponencial
vb2 <- vcov(ajust4) [3,3];vb2
vG <- grad %*% vb2 %*% grad;vG
sqrt(vG) # Std. Error do scale
#
#Calculo do AIC (-2*loglik+2*(p+k) em que p é o num de param da
# dist e k é o num de covariaveis
ajust4_AIC<- -2*ajust4$loglik[2]+2*(2+9);ajust4_AIC
#Calculo do BIC (-2*loglik+(p+k)*log(r) #r é o tamanho da amostra
r<-length(dados$Censura);r
ajust4_BIC<- -2*ajust4$loglik[2]+(2+9)*log(r);ajust4_BIC
#
#Calcular o Std. Error do exp(Log(scale))
vb <- vcov(ajust4);vb #ou ajust4$var
grad <- ajust4$scale;grad #ja em exponencial
vb2 <- vcov(ajust4) [3,3];vb2
vG <- grad %*% vb2 %*% grad;vG
sqrt(vG) # Std. Error do scale

# 5-Modelo Logistico
ajust5<-survreg(Surv(dados$Tempo/100, dados$Censura) ~
Unidades_de_conservacao+
                terras_indigenas+Pastagem+ Potencialidade.agrícola +
                Areas_urbanas + scale(distancia_estrada) +
                scale(distancia_hidrovia)+
                scale(distancia_mineração) + scale(População),
                dist = "logistic", data = dados)
summary(ajust5)
xb5<-ajust5$linear.predictors;xb5
sigma5<-ajust5$scale;sigma5
stlog <- 1/(1+exp((dados$Tempo/100-xb5)/sigma5));stlog
ajust5$loglik
AIC(ajust5)
BIC(ajust5)

#Calculo do AIC (-2*loglik+2*(p+k) em que p é o num de param da
# dist e k é o num de covariaveis
ajust5_AIC<- -2*ajust5$loglik[2]+2*(2+7);ajust5_AIC #Logistica
#Calculo do BIC (-2*loglik+(p+k)*log(r) em que p é o num de param da
# dist e r é o tamanho da amostra
r<-length(dados$Censura);r
ajust5_BIC<- -2*ajust5$loglik[2]+(2+7)*log(r);ajust5_BIC #Logistica

#Calcular o Std. Error do exp(Log(scale))
vb <- vcov(ajust5);vb #ou ajust5$var
grad <- ajust5$scale;grad #ja em exponencial
vb2 <- vcov(ajust5) [3,3];vb2
vG <- grad %*% vb2 %*% grad;vG
sqrt(vG) # Std. Error do scale

# 6-Modelo Gaussiano (Normal)
ajust6<-survreg(Surv(dados$Tempo/100, dados$Censura) ~
Unidades_de_conservacao+
                terras_indigenas+Pastagem+ Potencialidade.agrícola +
                Areas_urbanas + scale(distancia_estrada) +
                scale(distancia_hidrovia)+
                scale(distancia_mineração) + scale(População),
                dist = "gaussian", data = dados)
summary(ajust6)
xb6<-ajust6$linear.predictors;xb6

```

```

sigma6<-ajust6$scale;sigma6
stgau <- 1-pnorm(((dados$Tempo-xb6)/sigma6);stgau
ajust6$loglik
AIC(ajust6)
BIC(ajust6)

# Calculo do AIC
ajust6_AIC<- -2*ajust6$loglik[2]+2*(2+7); ajust6_AIC #Normal
#Calculo do BIC (-2*loglik+(p+k)*log(r) em que p é o num de param da
#dist e r é o número total de falhas
r<-length(dados$Censura);r
ajust6_BIC<- -2*ajust6$loglik[2]+(2+7)*log(r);ajust6_BIC

#Calcular o Std. Error do exp(Log(scale))
vb <- vcov(ajust6);vb #ou ajust4$var
grad <- ajust6$scale;grad #ja em exponencial
vb2 <- vcov(ajust6)[3,3];vb2
vG <- grad %*% vb2 %*% grad;vG
sqrt(vG) # Std. Error do scale

# Tabela de resultados
model <- c("Exponencial", "Weibull", "LogNormal",
           "Log-Logistico","Logistico","Gaussiano")
AIC <- as.numeric(c(AIC(ajust1),AIC(ajust2),AIC(ajust3),
                   AIC(ajust4),AIC(ajust5),AIC(ajust6)))
BIC <- as.numeric(c(BIC(ajust1),BIC(ajust2),BIC(ajust3),
                   BIC(ajust4),BIC(ajust5),BIC(ajust6)))
resultados <- cbind(AIC,BIC)
rownames(resultados) <- model
colnames(resultados) <- c("AIC","BIC")
print(resultados, digits = 7)
min(AIC)
min(BIC)

# AIC
AIC_results <-c(ajust1_AIC,ajust2_AIC,ajust3_AIC,
               ajust4_AIC,ajust5_AIC,ajust6_AIC)
AIC_results

# BIC
BIC_results <-c(ajust1_BIC,ajust2_BIC,ajust3_BIC,
               ajust4_BIC,ajust5_BIC,ajust6_BIC)
BIC_results

# Akaike weights
require(mvMORPH)
aicw(AIC_results)

# AICc weights para pequenas amostras
aicw(AIC_results, aicc=TRUE)

require(cogmod)
AICw(AIC_results)
BICw(BIC_results)

#CALCULO DOS RESIDUOS
#=====

# 1-Residuos do Modelo Exponencial
#Residuos de Cox-Snell

```

```

rcls1<- -log(ste);rcls1 # residuos cox-snell
ekmlcs<-survfit(Surv(rcls1,dados$Censura)~1,data=dados)
t1cs<-ekmlcs$time;t1cs
ste_km<-ekmlcs$surv;ste_km
Sexp_cs<-exp(-t1cs);Sexp_cs
#exponencial padrao
par(mfrow=c(1,2))
plot(ste_km,Sexp_cs,xlim=c(0,1),ylim=c(0,1),xlab="S(ei): Kaplan-Meier",
      ylab="S(ei): Exponencial padrao",pch=16)
lines(c(0,1), c(0,1), type="l", lty=3)
plot(ekmlcs,conf.int=F,mark.time=F, xlab="Cox-Snell residuals",
      ylab="Survival probability",xlim=c(0,3.0),main="Exponential")
lines(t1cs,Sexp_cs,lty=2)
legend(1.05, 0.5,lty=c(1,2),c("Kaplan-Meier","Standard Exponential"),
      cex=0.6,bty="n")

#=====

# 2-Residuos do Modelo Weibull
#Residuos de Cox-Snell
tiff("(TIFF) modeloparametricoCoxSnell (Ingles).tiff", width = 200,
      height = 120, units = "mm", res = 600, family = 'Calibri')
rcls2 <- -log(stw) # Cox-Snell residuals
ekm2cs <- survfit(Surv(rcls2, dados$Censura) ~ 1, data = dados)
t2cs <- ekm2cs$time
stw_km <- ekm2cs$surv
Stw_cs <- exp(-t2cs)

# Standard Weibull
par(mfrow = c(1, 2))

# First plot
plot(ekm2cs, conf.int = FALSE, mark.time = FALSE,
      xlab = "Cox-Snell Residuals",
      ylab = "Survival Probability",
      xlim = c(0, 4), main = "")
lines(t2cs, Stw_cs, lty = 2)
legend(1.4, 0.5, lty = c(1, 2),
      legend = c("Kaplan-Meier", "Standard Exponential"),
      cex = 0.6, bty = "n")
# Adding the title in the upper left corner
mtext("A", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

# Second plot
plot(stw_km, Stw_cs, xlim = c(0, 1), ylim = c(0, 1),
      xlab = "S(ei): Kaplan-Meier",
      ylab = "S(ei): Standard Exponential", pch = 16)
lines(c(0, 1), c(0, 1), type = "l", lty = 3)

# Adding the title in the upper left corner
mtext("B", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)
dev.off()

tiff("(TIFF) modeloparametricoCoxSnell.tiff", width = 200,
      height = 120, units = "mm", res = 600, family = 'Calibri')
rcls2 <- -log(stw) # Residuos de Cox-Snell
ekm2cs <- survfit(Surv(rcls2, dados$Censura) ~ 1, data = dados)
t2cs <- ekm2cs$time
stw_km <- ekm2cs$surv

```

```

Stw_cs <- exp(-t2cs)

# Weibull padrão
par(mfrow = c(1, 2))

# Primeiro gráfico
plot(ekm2cs, conf.int = FALSE, mark.time = FALSE,
     xlab = "Resíduos de Cox-Snell",
     ylab = "Probabilidade de Sobrevivência",
     xlim = c(0, 4), main = "")
lines(t2cs, Stw_cs, lty = 2)
legend(1.4, 0.5, lty = c(1, 2),
       legend = c("Kaplan-Meier", "Exponencial Padrão"),
       cex = 0.6, bty = "n")
# Adicionando o título na parte superior esquerda
mtext("A", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)
# Adicionando o título na parte superior esquerda

plot(stw_km, Stw_cs, xlim = c(0, 1), ylim = c(0, 1),
     xlab = "S(ei): Kaplan-Meier",
     ylab = "S(ei): Exponencial Padrão", pch = 16)
lines(c(0, 1), c(0, 1), type = "l", lty = 3)

# Segundo gráfico

mtext("B", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)
#
dev.off()

#=====
# 3-Resíduos do Modelo Log-normal
#Resíduos de Cox-Snell
rcs3<- -log(stln);rcs3 # residuos cox-snell
ekm3cs<-survfit(Surv(rcs3,dados$e)~1,data=dados)
t3cs<-ekm3cs$time;t3cs
stln_km<-ekm3cs$urv;stln_km
Stln_cs<-exp(-t3cs);Stln_cs

# Log-normal padrao
par(mfrow=c(1,2))
plot(stln_km,Stln_cs,xlim=c(0,1),ylim=c(0,1),xlab="S(ei): Kaplan-Meier",
     ylab="S(ei): Exponencial padrao",pch=16)
lines(c(0,1), c(0,1), type="l", lty=3)
plot(ekm3cs,conf.int=F,mark.time=F, xlab="Cox-Snell residuals",
     ylab="Survival probability",xlim=c(0,2.5),main="Log-normal")
lines(t3cs,Stln_cs,lty=2)
legend(1.4, 0.5,lty=c(1,2),c("Kaplan-Meier","Standard Exponential"),
       cex=0.6,bty="n")

#=====

# 4-Resíduos do Modelo Log-logístico
#Resíduos de Cox-Snell
rcs4<- -log(stllog);rcs4 # residuos cox-snell
ekm4cs<-survfit(Surv(rcs4,dados$e)~1,data=dados)
t4cs<-ekm4cs$time;t4cs
stllog_km<-ekm4cs$urv;stllog_km
Stllog_cs<-exp(-t4cs);Stllog_cs
# Log-logístico padrao
par(mfrow=c(1,2))
plot(stllog_km,Stllog_cs,xlim=c(0,1),ylim=c(0,1),xlab="S(ei): Kaplan-

```

```

Meier",
  ylab="S(ei): Exponencial padrao",pch=16)
lines(c(0,1), c(0,1), type="l", lty=3)
plot(ekm4cs,conf.int=F,mark.time=F, xlab="Cox-Snell residuals",
     ylab="Survival probability",xlim=c(0,2.5),main="Log-logistic")
lines(t4cs,Stllog_cs,lty=2)
legend(-0.1,0.2,lty=c(1,2),c("Kaplan-Meier","Standard Exponential"),
       cex=0.6,bty="n")

#=====

# 5-Residuos do Modelo Logistico
#Residuos de Cox-Snell
AIC
rcs5<- -log(stlog);rcs5 # residuos cox-snell
ekm5cs<-survfit(Surv(rcs5,dados$e)~1,data=dados)
t5cs<-ekm5cs$time;t5cs
min(ekm5cs$urv)
stlog_km<-ekm5cs$urv;stlog_km
Stlog_cs<-exp(-t5cs);Stlog_cs

# Graficos adequação do modelo
par(mfrow=c(1,2))
plot(stlog_km,Stlog_cs,xlim=c(0,1),ylim=c(0,1),xlab="S(ei): Kaplan-Meier",
     ylab="S(ei): Exponencial padrao",pch=16)
lines(c(0,1), c(0,1), type="l", lty=3)
plot(ekm5cs,conf.int=F,mark.time=F, xlab="Cox-Snell residuals",
     ylab="Survival probability",xlim=c(0,3),main="Logistic")
lines(t5cs,Stlog_cs,lty=2)
legend(-0.1,0.3,lty=c(1,2),c("Kaplan-Meier","Standard Exponential"),
       cex=0.6,bty="n")

#=====

# 6-Residuos do Modelo Gaussiano
#Residuos de Cox-Snell
rcs6<- -log(stgau);rcs6 # residuos cox-snell
ekm6cs<-survfit(Surv(rcs6,dados$e)~1,data=dados)
t6cs<-ekm6cs$time;t6cs
stgau_km<-ekm6cs$urv;stgau_km
Stgau_cs<-exp(-t6cs);Stgau_cs

# Gaussiano
par(mfrow=c(1,2))
plot(stgau_km,Stgau_cs,xlim=c(0,1),ylim=c(0,1),xlab="S(ei): Kaplan-Meier",
     ylab="S(ei): Exponencial padrao",pch=16)
lines(c(0,1), c(0,1), type="l", lty=3)
plot(ekm6cs,conf.int=F,mark.time=F, xlab="Cox-Snell residuals",
     ylab="Survival probability",xlim=c(0,3.0),main="Gaussian")
lines(t6cs,Stgau_cs,lty=2)
legend(-0.1,0.3,lty=c(1,2),c("Kaplan-Meier","Standard Exponential"),
       cex=0.6,bty="n")

```

APÊNDICE G – Script para o ajuste do modelo de fragilidade

```

# MODELO DE FRAGILIDADE

rm(list = ls())
library(INLA)
library(spdep)
library(survival)
#library(devtools)
#install_github("https://github.com/cran/maptools.git")
library(maptools)
library(ggplot2)
library(viridis)
library(tmap) #Pacote para fazer o grafico de mapa
library(readxl)
library(knitr)
library(ggpubr)
library(dplyr)
library(brinla)

#source("contagem_customizada.R") # A função organiza os termos de
#codigo de malha para partir de 1

#Lendo a base de dados
dados = read_excel("Base_Final.xlsx")

#Filtrando a base de dados para pegar somente o tipo de vegetação
#ombrofila densa
dados = dplyr::filter(dados, dados$Vegetação == "Floresta Ombrófila Densa")
table(dados$Vegetação)

#####

#Right join para match de códigos dos grupos com o codigo de indice do mapa

mapa=readShapePoly("camada_clusters.shp",
                  IDvar=NULL, proj4string=CRS(as.character(NA)),
                  verbose=FALSE, repair=FALSE, force_ring=FALSE,
                  delete_null_obj=FALSE,
                  retrieve_ABS_null=FALSE)

dados = subset(dados, dados$Tempo != 0)

matriz=poly2nb(mapa)
W=nb2mat(matriz, glist=NULL, style="B", zero.policy=NULL)
#rownames(W) = colnames(W) = contagem_customizada(indices)

adj_matrix = W
nrow(dados)
plot(mapa, border="black", axes=TRUE, las=1)

# Adicionar os rótulos aos polígonos
text(coordinates(mapa), labels=mapa$id, cex=0.6, pos=1)

# Converter a matriz de adjacência para um objeto INLA
adj <- inla.read.graph(adj_matrix)

# Preparar a variável de sobrevivência como duas colunas separadas:
#tempo e censura

```

```

dados$event <- dados$Censura # 1 para evento, 0 para censura
dados$event = as.numeric(dados$event)

# Formatando as variaveis
dados$Unidades_de_conservacao[is.na(dados$Unidades_de_conservacao)] <- 0
dados$terras_indigenas[is.na(dados$terras_indigenas)] = 0
dados$`Areas_urbanas`[is.na(dados$`Areas_urbanas`)] = 0
table(dados$`Areas_urbanas`)
summary(dados$distancia_estrada)
sd(dados$distancia_mineraçao)

#Definindo a formula do modelo
formula <- inla.surv(dados$Tempo/100, dados$Censura) ~
Unidades_de_conservacao+
  terras_indigenas+Pastagem+ Potencialidade.agrícola +
  scale(distancia_estrada) + scale(distancia_hidrovia)+
  scale(distancia_mineraçao) + scale(População) #+ f(cluster_id,
#model = "besag", graph = adj)

#Priori para os efeitos fixos
prior_fixed <- list(
  mean = list(rep(0, 8)),
  prec = list(rep(0.001, 8)
)
)

# Priori para a precisão do efeito aleatório
control.hyper <- list(
  hyper = list(
    prec = list(
      prior = "loggamma", # Priori Log-Gamma para a precisão
      param = c(1, 100000000) # Parâmetros da Log-Gamma: forma = 1, taxa =
0.01
    )
  )
)

#Ajustando o modelo
modelo_weibull <- inla(
  formula = formula,
  data = dados,
  family = "weibullsurv",
  control.fixed = list(
    mean = unlist(prior_fixed$mean),
    prec = unlist(prior_fixed$prec)
  ),
  control.compute = list(dic = TRUE, waic = TRUE, cpo = TRUE, config =
TRUE),
  control.family = list(variant = 1),
  verbose = F
)

summary(modelo_weibull)
round(modelo_weibull$summary.fixed, 4)

#Ajustando o modelo

```

```

modelo_weibull_frailty <- inla(
  formula = formula,
  data = dados,
  family = "weibullsurv",
  control.fixed = list(
    mean = unlist(prior_fixed$mean),
    prec = unlist(prior_fixed$prec)
  ),
  control.compute = list(dic = TRUE, waic = TRUE, cpo = TRUE, config =
TRUE),
  control.family = list(variant = 1),
  verbose = TRUE
)

modelo_exponencial <- inla(
  formula = formula,
  data = dados,
  family = "exponentialsurv",
  control.fixed = list(
    mean = prior_fixed$mean,
    prec = prior_fixed$prec
  ),
  control.compute = list(dic = TRUE, waic = TRUE, cpo = TRUE, config =
TRUE),
  verbose = TRUE
)
summary(modelo_exponencial)

modelo_lognormal <- inla(
  formula = formula,
  data = dados,
  family = "lognormalsurv",
  control.fixed = list(
    mean = prior_fixed$mean,
    prec = prior_fixed$prec
  ),
  control.compute = list(dic = TRUE, waic = TRUE, cpo = TRUE, config =
TRUE),
  verbose = F
)

summary(modelo_lognormal)

# FUNCAO PARA O PLOT DO GRÁFICO
plot_coxsnell_res <- function(resid.obj, lwd = 1.6, xlab="Resíduos de Cox
Sell",
                             ylab="Estimativa do risco acumulado",
                             main = "", cex.lab = 1, cex.axis = 0.8){
  require(survival)
  cs.resid <- resid.obj$cs
  event <- resid.obj$event
  s.cs.res <- survfit(Surv(cs.resid, event) ~ 1, type="fleming-harrington")
  H.est <- cumsum(s.cs.res$n.event/s.cs.res$n.risk)
  xy.max <- max(c(s.cs.res$time, H.est))
  plot(s.cs.res$time, H.est, type='s', col='black', xlab=xlab,
       ylab=ylab, lwd=lwd, xlim=c(0,xy.max), ylim=c(0,xy.max),
       main = main, cex.lab = cex.lab, cex.axis = cex.axis)
  abline(0, 1, lty=2, col = "red")
}

```

```

tiff("(TIFF) residuoCoxSnellINLA_corrigido_portugues.tiff",width=200,
      height=150,units="mm",res=600,family='Calibri')

#PLOTANDO OS GRÁFICOS
layout.m <- matrix(c(rep(1:2, each=2),
                      0, rep(3, each=2), 0),
                  nrow=2, byrow=T)

layout(layout.m)

# Ajustando as margens para que todos os gráficos tenham tamanhos similares
par(mar=c(4, 4.5, 3, 1)) # Mantendo proporções equilibradas

#Resíduos para o modelo exponencial
res = bri.surv.resid(modelo_exponencial, dados$Tempo/100, dados$Censura)
#bri.csresid.plot(res)
plot_coxsnell_res(res)
#title(main = "Exponencial") # Adicionar título ao gráfico
# Adicionando o título na parte superior esquerda
mtext("A", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

#O grafico do modelo weibull so é plotado na parametrização variant = 1

#Resíduos para o modelo Weibull
res = bri.surv.resid(modelo_weibull, dados$Tempo/100, dados$Censura)
#bri.csresid.plot(res)
plot_coxsnell_res(res)
#title(main = "Weibull") # Adicionar título ao gráfico

# Adicionando o título na parte superior esquerda
mtext("B", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

# AVALIAÇÃO DO MODELO (LOGNORMAL)

summary(modelo_lognormal)
# Obter o preditor linear (mu)
theta <- modelo_lognormal$summary.fixed[-1,1]
mu <- modelo_lognormal$summary.linear.predictor$mean; mu
# theta são os coeficientes estimados

modelo_lognormal$summary.hyperpar
# Obter a escala (sigma)
sigma <- 1 / sqrt(1.59) # Baseado na precisão estimada pelo INLA

# Função de sobrevivência acumulada S(t)
S_t <- 1 - pnorm((log(dados$Tempo/100) - mu) / sigma); S_t

# Resíduos de Cox-Snell
cs.resid <- -log(S_t)
min(S_t)

#Resíduos de Cox-Snell
r3cs <- -log(S_t); r3cs # residuos cox-snell
ekm3cs <- survfit(Surv(r3cs, dados$Censura) ~ 1, data = dados)
t3cs <- ekm3cs$time; t3cs
stln_km <- ekm3cs$surv; stln_km
Stln_cs <- exp(-t3cs); Stln_cs

s.cs.res <- survfit(Surv(r3cs, dados$Censura) ~ 1, data = dados,

```

```

                                type="fleming-harrington")
H.est <- cumsum(s.cs.res$n.event/s.cs.res$n.risk)
xy.max <- max(c(s.cs.res$time, H.est))
plot(s.cs.res$time, H.est,type='s',col='black', lwd = 1.6,
      xlab="Resíduos de Cox Snell", ylab="Estimativa do risco acumulado",
      main = "", cex.lab = 1, cex.axis = 0.8)
abline(0, 1, lty=2, col = "red")

# Adicionando o título na parte superior esquerda
mtext("C", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)

#Resíduos para o modelo cox
res = bri.surv.resid(modelo_cox, dados$Tempo/100, dados$Censura)
#bri.csresid.plot(res)
plot_coxsnell_res(res)
#title(main = "Cox") # Adicionar título ao gráfico
# Adicionando o título na parte superior esquerda
mtext("D", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)
#save.image("Meu_ambiente_1-Modelos de fragilidade.RData")
dev.off()

tiff("(TIFF) residuoCoxSnellINLA_PORTUGUÊS.tiff",width=200,
      height=150,units="mm",res=600,family='Calibri')
#PLOTANDO OS GRÁFICOS

layout.m <- matrix(c(rep(1:2, each=2),
                      0, rep(3, each=2), 0),
                  nrow=2, byrow=T)

layout(layout.m)

# Ajustando as margens para que todos os gráficos tenham tamanhos similares
par(mar=c(4, 4.5, 3, 1)) # Mantendo proporções equilibradas

# Resíduos para o modelo Exponencial
res = bri.surv.resid(modelo_exponencial, dados$Tempo/100, dados$Censura)
plot_coxsnell_res(res)
mtext("A", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)
#mtext("Exponencial", side = 3, line = -1, cex = 0.9)

# Resíduos para o modelo Weibull (apenas na parametrização variant = 1)
res = bri.surv.resid(modelo_weibull, dados$Tempo/100, dados$Censura)
plot_coxsnell_res(res)
mtext("B", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)
#mtext("Weibull", side = 3, line = -1, cex = 0.9)

# Avaliação do modelo Lognormal
summary(modelo_lognormal)

# Obter o preditor linear (mu)
theta <- modelo_lognormal$summary.fixed[-1,1]
mu <- modelo_lognormal$summary.linear.predictor$mean

# Obter a escala (sigma)
sigma <- 1 / sqrt(1.59)

# Função de sobrevivência acumulada S(t)
S_t <- 1 - pnorm((log(dados$Tempo/100) - mu) / sigma)

# Resíduos de Cox-Snell
cs.resid <- -log(S_t)

```

```

# Cálculo dos resíduos
rcs3 <- -log(S_t)
ekm3cs <- survfit(Surv(rcs3, dados$Censura) ~ 1, data = dados)
t3cs <- ekm3cs$time
stln_km <- ekm3cs$surv
Stln_cs <- exp(-t3cs)

s.cs.res <- survfit(Surv(rcs3, dados$Censura) ~ 1, data = dados,
                  type = "fleming-harrington")
H.est <- cumsum(s.cs.res$n.event / s.cs.res$n.risk)
xy.max <- max(c(s.cs.res$time, H.est))

plot(s.cs.res$time, H.est, type = 's', col = 'black', lwd = 1.6,
      xlab = "Resíduos de Cox-Snell",
      ylab = "Risco Acumulado Estimado",
      main = "", cex.lab = 0.65, cex.axis = 0.65)
abline(0, 1, lty = 2, col = "red")

mtext("C", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)
#mtext("Lognormal", side = 3, line = -1, cex = 0.9)

# Resíduos para o modelo Cox
res = bri.surv.resid(modelo_cox, dados$Tempo/100, dados$Censura)
plot_coxsnell_res(res)
mtext("D", side = 3, line = 0.5, adj = 0, font = 1, cex = 0.8)
#mtext("Cox", side = 3, line = -1, cex = 0.9)

dev.off()

#####
pal <- c("#00008FFF", "#0000F2FF", "#0063FFFF", "#00D4FFFF", "#46FFB8FF",
        "#B8FF46FF", "#FFD400FF", "#FF6300FF", "#F00000FF", "#800000FF")

custom_theme <- theme_bw() + theme(legend.position = "right",
                                   text = element_text(size = 14),
                                   plot.title = element_text(size = 16),
                                   legend.title = element_text(size = 12))

# Criando o grafico das marginais

post_fitted <- modelo_weibull$marginals.fixed
summary(modelo_weibull)
post_margin <- data.frame(do.call(rbind, post_fitted))
post_margin$cement <- rep(names(post_fitted),
                          times = sapply(post_fitted, nrow))
post_margin$xTransformed = -post_margin$x*0.3759398

tiff("(TIFF) marginaisParametrosComEixo.tiff",width=200,height=150,
     units="mm",res=600,family='Calibri')

ggplot(post_margin) +
  aes(x = x, y = y, colour = cement, linetype = cement, group = cement) +
  geom_line(linewidth = 0.95) +
  geom_hline(yintercept = 0, colour = "black") + # Linha do eixo X
  geom_vline(xintercept = 0, colour = "black") + # Linha do eixo Y
  scale_color_manual(
    values = c(`(Intercept)` = "#FF1606",
              terras_indigenas = "#DD8A1D",
              Unidades_de_conservacao = "#B59C00",

```

```

    Pastagem = "#78AC0A",
    Potencialidade.agricolaB = "#0DB832",
    Potencialidade.agricolaC = "#00BE79",
    `Potencialidade.agricolaCorpo d'água continental` = "blue",
    Potencialidade.agricolaD = "#08B6E5",
    `scale(distancia_estrada)` = "#4FA1F9",
    `scale(distancia_hidrovia)` = "#A385FC",
    `scale(distancia_mineração)` = "brown",
    `scale(População)` = "#FF61C3"),
  labels = c(
    `(Intercept)` = "Intercepto",
    Unidades_de_conservacao = "UC - Sim",
    terras_indigenas = "Terras indígenas - Sim",
    Pastagem = "Pastagem",
    Potencialidade.agricolaB = "PA - Boa",
    Potencialidade.agricolaC = "PA - Moderada",
    `Potencialidade.agricolaCorpo d'água continental` = "PA - Corpo
d'água",
    Potencialidade.agricolaD = "PA - Fort. Restrita",
    `scale(distancia_estrada)` = "Distância às estradas",
    `scale(distancia_hidrovia)` = "Distância às hidrovias",
    `scale(distancia_mineração)` = "Distância à mineração",
    `scale(População)` = "População municipal"
  )
) +
scale_linetype_manual(
  values = c(
    `(Intercept)` = "solid",
    terras_indigenas = "dashed",
    Unidades_de_conservacao = "dotted",
    Pastagem = "dotdash",
    Potencialidade.agricolaB = "twodash",
    Potencialidade.agricolaC = "longdash",
    `Potencialidade.agricolaCorpo d'água continental` = "solid",
    Potencialidade.agricolaD = "dashed",
    `scale(distancia_estrada)` = "dotted",
    `scale(distancia_hidrovia)` = "dotdash",
    `scale(distancia_mineração)` = "twodash",
    `scale(População)` = "longdash"
  )
) +
labs(x = "Coeficientes", color = "Variável", linetype = "Variável") +
guides(
  color = guide_legend(override.aes = list(linetype = c(
    "solid", "dotdash", "twodash", "longdash", "solid", "dashed",
    "dotted", "dotdash", "twodash", "longdash", "dashed", "dotted"
  ))),
  linetype = "none" # Remove a legenda de linetype
) +
theme_minimal() +
theme(
  panel.grid.major = element_blank(),
# Remove as linhas de grade maiores
  panel.grid.minor = element_blank()
# Remove as linhas de grade menores
)

dev.off()

tiff("(TIFF) marginalsParametersSemEixo (Ingles).tiff", width = 200,

```

```

height = 150, units = "mm", res = 600, family = 'Calibri')

ggplot(post_margin) +
  aes(x = x, y = y, colour = cement, linetype = cement, group = cement) +
  geom_line(linewidth = 0.95) +
  #geom_hline(yintercept = 0, colour = "black") + # Linha do eixo X+
  #geom_vline(xintercept = 0, colour = "black") + # Linha do eixo Y+
  scale_color_manual(
    values = c(`(Intercept)` = "#FF1606",
              terras_indigenas = "#DD8A1D",
              Unidades_de_conservacao = "#B59C00",
              Pastagem = "#78AC0A",
              Potencialidade.agricolaB = "#0DB832",
              Potencialidade.agricolaC = "#00BE79",
              `Potencialidade.agricolaCorpo d'água continental` = "blue",
              Potencialidade.agricolaD = "#08B6E5",
              `scale(distancia_estrada)` = "#4FA1F9",
              `scale(distancia_hidrovia)` = "#A385FC",
              `scale(distancia_mineração)` = "brown",
              `scale(População)` = "#FF61C3"),
    labels = c(
      `(Intercept)` = "Intercept",
      Unidades_de_conservacao = "CU - Yes",
      terras_indigenas = "Indigenous land - Yes",
      Pastagem = "Pasture",
      Potencialidade.agricolaB = "AP - Good",
      Potencialidade.agricolaC = "AP - Moderate",
      `Potencialidade.agricolaCorpo d'água continental` = "AP - Water",
      Potencialidade.agricolaD = "AP - Highly restricted",
      `scale(distancia_estrada)` = "Distance to roads",
      `scale(distancia_hidrovia)` = "Distance to waterways",
      `scale(distancia_mineração)` = "Distance to mining",
      `scale(População)` = "Municipal population"
    )
  ) +
  scale_linetype_manual(
    values = c(
      `(Intercept)` = "solid",
      terras_indigenas = "dashed",
      Unidades_de_conservacao = "dotted",
      Pastagem = "dotdash",
      Potencialidade.agricolaB = "twodash",
      Potencialidade.agricolaC = "longdash",
      `Potencialidade.agricolaCorpo d'água continental` = "solid",
      Potencialidade.agricolaD = "dashed",
      `scale(distancia_estrada)` = "dotted",
      `scale(distancia_hidrovia)` = "dotdash",
      `scale(distancia_mineração)` = "twodash",
      `scale(População)` = "longdash"
    )
  ) +
  labs(x = "Coefficients", color = "Variable", linetype = "Variable") +
  guides(
    color = guide_legend(override.aes = list(linetype = c(
      "solid", "dotdash", "twodash", "longdash", "solid", "dashed",
      "dotted", "dotdash", "twodash", "longdash", "dashed", "dotted"
    ))),
    linetype = "none" # Remove linetype legend
  ) +
  theme_minimal() +
  theme(

```

```

    panel.grid.major = element_blank(), # Remove major grid lines
    panel.grid.minor = element_blank() # Remove minor grid lines
  )
dev.off()

tiff("(TIFF) distribuicaomarginalHIperparametro.tiff", width = 200,
     height = 150, units = "mm", res = 600, family = 'Calibri')

ggplot(post_margin) +
  aes(x = x, y = y, colour = cement) +
  # Definindo a cor conforme o parâmetro
  geom_line() + # Usando geom_line() para curvas
  scale_colour_manual(values = c("alpha parameter for weibullsurv" = "red",
                                "Precision for cluster_id" = "blue"),
                     labels = c("Alpha", "Parâmetro de precisão")) +
  # Alterando os rótulos da legenda
  theme_minimal() +
  facet_wrap(vars(cement), scales = "free",
            labeller = as_labeller(c("alpha parameter for weibullsurv" =
"A",
                                "Precision for cluster_id" = "B")))) +

  # Títulos (a) e (b)
  labs(
    x = "Distribuição dos hiperparâmetros", # Nome do eixo X
    colour = "Parâmetro" # Título da legenda
  ) +
  theme(
    legend.position = c(0.85, 0.5),
    # Colocando a legenda dentro do gráfico 2, ajustando a posição
    strip.background = element_blank(),
    # Remover o fundo da faixa de faceta
    strip.text = element_text(hjust = 0, vjust = 1, size = 12),
    # Alinhamento do título da faceta no canto superior esquerdo
    axis.title.x = element_text(size = 12), # Adiciona título do eixo X
    axis.title.y = element_text(size = 12), # Adiciona título do eixo Y
    axis.text.x = element_text(size = 10), # Tamanho do texto do eixo X
    axis.text.y = element_text(size = 10), # Tamanho do texto do eixo Y
    panel.grid.major = element_blank(), # Remove as linhas de grade
    maiores
    panel.grid.minor = element_blank(), # Remove as linhas de grade
    menores
    axis.line = element_line(colour = "black") # Adiciona linhas de eixo
  )
dev.off()

tiff("(TIFF) distribuicaomarginalHIperparametro(Ingles).tiff", width = 200,
     height = 150, units = "mm", res = 600, family = 'Calibri')
ggplot(post_margin) +
  aes(x = x, y = y, colour = cement) +
  # Definindo a cor conforme o parâmetro
  geom_line() + # Usando geom_line() para curvas
  scale_colour_manual(values = c("alpha parameter for weibullsurv" = "red",
                                "Precision for cluster_id" = "blue"),
                     labels = c("Alpha", "Precision parameter")) +
  # Alterando os rótulos da legenda
  theme_minimal() +
  facet_wrap(vars(cement), scales = "free",

```

```

    labeller = as_labeller(c("alpha parameter for weibullsurv" =
"A",
                                "Precision for cluster_id" = "B")) +
# Títulos (a) e (b)
labs(
  x = "Distribution of hyperparameters", # Nome do eixo X
  colour = "Parameter" # Título da legenda
) +
theme(
  legend.position = c(0.85, 0.5),
  # Colocando a legenda dentro do gráfico 2, ajustando a posição
  strip.background = element_blank(),
  # Remover o fundo da faixa de faceta
  strip.text = element_text(hjust = 0, vjust = 1, size = 12),
  # Alinhamento do título da faceta no canto superior esquerdo
  axis.title.x = element_text(size = 12), # Adiciona título do eixo X
  axis.title.y = element_text(size = 12), # Adiciona título do eixo Y
  axis.text.x = element_text(size = 10), # Tamanho do texto do eixo X
  axis.text.y = element_text(size = 10), # Tamanho do texto do eixo Y
  panel.grid.major = element_blank(), # Remove as linhas de grade
maiores
  panel.grid.minor = element_blank(), # Remove as linhas de grade
menores
  axis.line = element_line(colour = "black") # Adiciona linhas de eixo
)
dev.off()

```